

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 September 2003 (25.09.2003)

PCT

(10) International Publication Number
WO 03/078637 A2

(51) International Patent Classification⁷: **C12N 15/62**,
15/12, 15/29, 15/54, 9/10, 15/10, A01H 5/00

Antonius [NL/NL]; Van Doesburglaan 160, NL-6708
MD Wageningen (NL). **BOSCH, Hendrik, Jan** [NL/NL];
Wim Sonneveldstraat 17, NL-6708 NA Wageningen (NL).
ROUWENDAL, Gerard, Johan, Adolph [NL/NL]; De
Rauwendaal 50, NL-6666 CE Heteren (NL).

(21) International Application Number: PCT/IB03/01626

(74) Agents: **MARSHALL, Cameron, J.** et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(22) International Filing Date: 18 March 2003 (18.03.2003)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(25) Filing Language: English

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,

(26) Publication Language: English

(30) Priority Data:
60/365,735 19 March 2002 (19.03.2002) US

(71) Applicant (*for all designated States except US*): **PLANT RESEARCH INTERNATIONAL B.V.** [NL/NL];
Droevendaalsesteeg 1, NL-6708 PB Wageningen (NL).

(72) Inventors; and

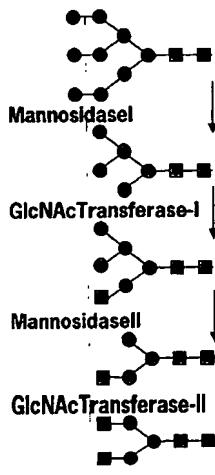
(75) Inventors/Applicants (*for US only*): **BAKKER, Hendrikus, Antonius, Cornelius** [DE/DE]; Drachenfeld 75, 30627 Hanover (DE). **FLORACK, Dionisius, Elisabeth**,

[Continued on next page]

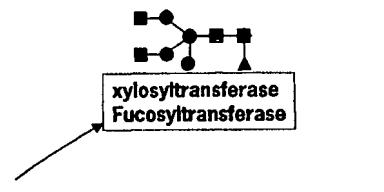
(54) Title: OPTIMIZING GLYCAN PROCESSING IN PLANTS

WO 03/078637 A2

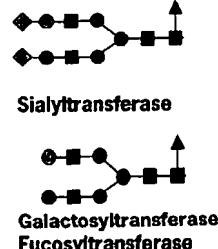
Common



Plants



Mammals



■ N-acetylglucosamine ● Mannose
● Galactose ▲ Fucose
○ Xylose ◆ Sialic acid

(57) Abstract: The invention is directed to methods for optimizing glycan processing in organisms (and in particular, plants) so that a glycoprotein having complex type bi-antennary glycans and thus containing galactose residues on both arms and which are devoid of (or reduce in) xylose and fucose can be obtained. The invention is further directed to said glycoprotein obtained and host system comprising said protein.



SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *without international search report and to be republished upon receipt of that report*

OPTIMIZING GLYCAN PROCESSING IN PLANTS

FIELD OF THE INVENTION

The invention is directed to methods for optimizing glycan processing of cell or an organism containing glycoproteins with N-glycans, in particular plants so that a glycoprotein having an N-glycan, high mannose type, hybrid or preferably complex type N-glycans, including but not limited to bi-antennary N-glycans, and containing a galactose residue on at least one arm of the N-glycan and which are devoid of (or reduced in) xylose and fucose residues can be obtained. The invention is further directed to said glycoprotein obtained and in particular a plant host system comprising said protein.

BACKGROUND OF THE INVENTION

N-linked glycans, specific oligosaccharide structures attached to asparagine residues of glycoproteins, can contribute significantly to the properties of the protein and, in turn, to the properties of the organism. Plant proteins can carry N-linked glycans but in marked contrast to mammals only few biological processes are known to which they contribute.

Biogenesis of N-linked glycans begins with the synthesis of a lipid linked oligosaccharide moiety (Glc3Man9GlcNAc2-) which is transferred en bloc to the nascent polypeptide chain in the endoplasmic reticulum (ER). Through a series of trimming reactions by exoglycosidases in the ER and cis-Golgi compartments, the so-called "high mannose" (Man9GlcNAc2 to Man5GlcNAc2) glycans are formed. Subsequently, the formation of complex type glycans starts with the transfer of the first GlcNAc onto Man5GlcNAc2 by GnTI and further trimming by mannosidase II (ManII) to form GlcNAcMan3GlcNAc2. Complex glycan biosynthesis continues while the glycoprotein is progressing through the secretory pathway with the transfer in the Golgi apparatus of the second GlcNAc residue by GnTII as well as other monosaccharide residues onto the GlcNAcMan3GlcNAc2 under the action of several other glycosyl transferases.

Plants and mammals differ with respect to the formation of complex glycans (see Figure 1, which compares the glycosylation pathway of glycoproteins in plants and mammals). In plants, complex glycans are characterized by the presence of $\beta(1,2)$ -xylose residues linked to the Man-3 and/or an $\alpha(1,3)$ -fucose residue linked to GlcNAc-1, instead of an $\alpha(1,6)$ -fucose residue linked to the GlcNAc-1. Genes encoding the corresponding xylosyl (XylT) and fucosyl (Fuct) transferases have been isolated [Strasser *et al.*, "Molecular cloning and functional expression of beta1,2-xylosyltransferase cDNA from *Arabidopsis thaliana*," *FEBS Lett.* 472:105 (2000); Leiter *et al.*, "Purification, cDNA cloning, and expression of GDP-L-Fuc:Asn-linked GlcNAc alpha 1,3-fucosyltransferase from mung beans," *J. Biol. Chem.* 274:21830 (1999)]. Plants do not possess $\beta(1,4)$ -galactosyltransferases nor $\alpha(2,6)$ sialyltransferases and consequently plant glycans lack the $\beta(1,4)$ -galactose and terminal $\alpha(2,6)$ NeuAc residues often found on mammalian glycans.

The final glycan structures are not only determined by the mere presence of enzymes involved in their biosynthesis and transport but to a large extent by the specific sequence of the various enzymatic reactions. The latter is controlled by discrete sequestering and relative position of these enzymes throughout the ER and Golgi, which is mediated by the interaction of determinants of the transferase and specific characteristics of the sub-Golgi compartment for which the transferase is destined. A number of studies using hybrid molecules have identified that the transmembrane domains of several glycosyltransferases, including that of $\beta(1,4)$ -galactosyltransferases, play a central role in their sub-Golgi sorting [Grabenhorst *et al.*, *J. Biol. Chem.* 274:36107 (1999); Colley, K., *Glycobiology* 7:1 (1997); Munro, S., *Trends Cell Biol.* 8:11 (1998); Gleeson, P.A., *Histochem. Cell Biol.* 109:517 (1998)].

Although plants and mammals have diverged a relatively long time ago, N-linked glycosylation seems at least partly conserved. This is evidenced by the similar though not identical glycan structures and by the observation that a mammalian GlcNAcTI gene complements a *Arabidopsis* mutant that is deficient in GlcNAcTI activity, and vice versa. The differences in glycan structures can have important consequences. For example, xylose and $\alpha(1,3)$ -fucose epitopes are known to be highly immunogenic and possibly allergenic in some circumstances, which may pose a problem when plants are used for the production of therapeutic glycoproteins. Moreover, blood serum of many allergy patients contains IgE directed against these epitopes but also 50% of non-allergic blood donors contains in their sera antibodies specific for core-xylose whereas 25% have antibodies for core-alpha 1,3-fucose (Bardor *et al.*, 2002, *in press, Glycobiology*) (Advance Access published December 17, 2002) which make these individuals at risk to treatments with recombinant proteins produced in plants containing fucose and/or xylose. In addition, this carbohydrate directed IgE in sera might cause false positive reaction in *in vitro* tests using plant extracts since there is evidence that these carbohydrate specific IgE's are not relevant for the allergenic reaction. In sum, a therapeutic failure with a glycoprotein produced in plants might be the result of accelerated clearance of the recombinant glycoprotein having xylose and/or fucose.

Accordingly, there is a need to better control glycosylation in plants, and particularly, glycosylation of glycoproteins intended for therapeutic use.

30 DEFINITIONS

To facilitate understanding of the invention, a number of terms as used in this specification are defined below.

The term "vector" refers to any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, retrovirus, virion, or similar genetic element, which is capable of replication when associated with the proper control elements and which can transfer gene sequences into cells and/or between cells. Thus, this term includes cloning and expression vehicles, as well as viral vectors.

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence (or coding sequences) – such as the coding sequence(s) for the hybrid enzyme(s) described in more detail below - and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host cell or organism.

5 Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals. It is not intended that the present invention be limited to particular expression vectors or expression vectors with particular elements.

10 The term "transgenic" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a cell, tissue or to a plant refers to a cell, tissue or plant, respectively, which comprises a transgene, where one or more cells of the tissue contain a transgene (such as a gene encoding the hybrid enzyme(s) of the present invention), or a plant whose genome 15 has been altered by the introduction of a transgene. Transgenic cells, tissues and plants may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

20 The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence" (*i.e.*, "foreign DNA"). The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (*e.g.*, a point mutation, the presence of a selectable marker gene, or other like modifications) relative to the naturally-occurring sequence.

25 The term "heterologous DNA sequence" refers to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell 30 into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (*e.g.*, proteins which confer drug resistance), or other similar elements.

35 The term "foreign gene" refers to any nucleic acid (*e.g.*, gene sequence) which is introduced into the genome of a cell by experimental manipulations and may include gene sequences found in that cell so long as the introduced gene contains some modification (*e.g.*, a point mutation, the presence of a selectable marker gene, or other like modifications) relative to the naturally-occurring gene.

The term "fusion protein" refers to a protein wherein at least one part or portion is from a first protein and another part or portion is from a second protein. The term "hybrid enzyme" refers to a fusion protein which is a functional enzyme, wherein at least one part or portion is from a first species and another part or portion is from a second species. Preferred hybrid enzymes of the 5 present invention are functional glycosyltransferases (or portions thereof) wherein at least one part or portion is from a plant and another part or portion is from a mammal (such as human).

The term "introduction into a cell" or "introduction into a host cell" in the context of nucleic acid (e.g., vectors) is intended to include what the art calls "transformation" or "transfection" or "transduction." Transformation of a cell may be stable or transient – and the present invention 10 contemplates introduction of vectors under conditions where, on the one hand, there is stable expression, and on the other hand, where there is only transient expression. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) 15 which detects the presence of a polypeptide encoded by one or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein (e.g., antigen binding of an antibody) encoded by the transgene (e.g., the antibody gene). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of 20 one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction (PCR) of genomic DNA of the cell to amplify 25 transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

The term "host cell" includes both mammalian (e.g. human B cell clones, Chinese hamster ovary cells, hepatocytes) and non-mammalian cells (e.g. insect cells, bacterial cells, plant cells). In 30 one embodiment, the host cells are mammalian cells and the introduction of a vector expressing a hybrid protein of the present invention (e.g TmGnTII-GalT) inhibits (or at least reduces) fucosylation in said mammalian cells.

The term "nucleotide sequence of interest" refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason (e.g., confer improved qualities, use 35 for production of therapeutic proteins), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (e.g., reporter genes, selection marker genes, oncogenes, antibody genes, drug resistance genes, growth factors, and other like genes), and non-coding regulatory sequences which do not encode an mRNA or protein product,

(e.g., promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, and other like sequences). The present invention contemplates host cells expressing a heterologous protein encoded by a nucleotide sequence of interest along with one or more hybrid enzymes.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from one or more other components (e.g., separated from a cell containing the nucleic acid, or separated from at least one contaminant nucleic acid, or separated from one or more proteins, one or more lipids) with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising SEQ ID NO:1 includes, by way of example, such nucleic acid sequences in cells which ordinarily contain SEQ ID NO:1 where the nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain at a minimum at least a portion of the sense or coding strand (i.e., the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and anti-sense strands (i.e., the nucleic acid sequence may be double-stranded).

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An "isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free, from other components with which they are naturally associated. The present invention contemplates both purified (including substantially purified) and unpurified hybrid enzyme(s) (which are described in more detail below).

As used herein, the terms "complementary" or "complementarity" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence. For example, the present invention contemplates the complements of SEQ ID NOS: 1, 3, 5, 9, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 40, 41 and 43.

5. The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (*i.e.*, partial identity) or complete homology (*i.e.*, complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the
10 completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (*i.e.*, an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous sequence to a target under conditions of low
15 stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-
20 complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

- 25 When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." [Coombs J (1994) *Dictionary of Biotechnology*, Stockton Press, New York NY]. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

- As used herein, the term " T_m " is used in reference to the "melting temperature." The melting
35 temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution

at 1 M NaCl [see e.g., Anderson and Young, Quantitative Filter Hybridization, in: *Nucleic Acid Hybridization* (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m .

Low stringency conditions when used in reference to nucleic acid hybridization comprise 5 conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE (Saline, Sodium Phosphate, EDTA) (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄•H₂O and 1.85 g/l EDTA (Ethylenediaminetetraacetic Acid), pH adjusted to 7.4 with NaOH), 0.1% SDS (Sodium dodecyl sulfate), 5X Denhardt's reagent [50X Denhardt's contains the following per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Bovine Serum Albumin) (Fraction V; Sigma)] and 100 µg/ml denatured 10 salmon sperm DNA followed by washing in a solution comprising between 0.2X and 2.0X SSPE, and 0.1% SDS at room temperature when a DNA probe of about 100 to about 1000 nucleotides in length is employed.

High stringency conditions when used in reference to nucleic acid hybridization comprise 15 conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE, 1% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, and 0.1% SDS at 68°C when a probe of about 100 to about 1000 nucleotides in length is employed.

The term "equivalent" when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization 20 condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results 25 in hybridization of the first nucleic acid sequence with the other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous 30 equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above-listed conditions.

35 The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (i.e., upstream) of a nucleotide sequence of interest whose transcription into

mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription.

Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide

- 5 sequence of interest to a specific type of tissue (e.g., petals) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (e.g., roots). Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of a plant such that the reporter construct is integrated into every tissue of the resulting transgenic plant,
- 10 and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic plant. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected. The term "cell type specific" as applied to a
- 15 promoter refers to a promoter which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell type
- 20 specificity of a promoter may be assessed using methods well known in the art, e.g., immuno-histochemical staining. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is controlled by the promoter. A labeled (e.g., peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to
- 25 bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (e.g., heat shock, chemicals, light, or similar stimuli). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue. In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (e.g., heat shock, chemicals, light, or similar stimuli) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

- The terms "infecting" and "infection" with a bacterium refer to co-incubation of a target
- 35 biological sample, (e.g., cell, tissue, plant part) with the bacterium under conditions such that nucleic acid sequences contained within the bacterium are introduced into one or more cells of the target biological sample.

The term "*Agrobacterium*" refers to a soil-borne, Gram-negative, rod-shaped phytopathogenic bacterium which causes crown gall. The term "*Agrobacterium*" includes, but is not limited to, the strains *Agrobacterium tumefaciens*, (which typically causes crown gall in infected plants), and *Agrobacterium rhizogens* (which causes hairy root disease in infected host plants).

- 5 Infection of a plant cell with *Agrobacterium* generally results in the production of opines (e.g., nopaline, agropine, octopine) by the infected cell. Thus, *Agrobacterium* strains which cause production of nopaline (e.g., strain LBA4301, C58, A208) are referred to as "nopaline-type" *Agrobacteria*; *Agrobacterium* strains which cause production of octopine (e.g., strain LBA4404, Ach5, B6) are referred to as "octopine-type" *Agrobacteria*; and *Agrobacterium* strains which cause 10 production of agropine (e.g., strain EHA105, EHA101, A281) are referred to as "agropine-type" *Agrobacteria*.

The terms "bombarding," "bombardment," and "biolistic bombardment" refer to the process of accelerating particles towards a target biological sample (e.g., cell, tissue, plant part – such as a leaf, or intact plant) to effect wounding of the cell membrane of a cell in the target biological sample 15 and/or entry of the particles into the target biological sample. Methods for biolistic bombardment are known in the art (e.g., U.S. Patent Nos. 5,584,807 and 5,141,131, the contents of both are herein incorporated by reference), and are commercially available (e.g., the helium gas-driven microparticle accelerator (PDS-1000/He) (BioRad)).

The term "microwounding" when made in reference to plant tissue refers to the introduction 20 of microscopic wounds in that tissue. Microwounding may be achieved by, for example, particle bombardment as described herein. The present invention specifically contemplates schemes for introducing nucleic acid which employ microwounding.

The term "organism" as used herein refers to all organisms and in particular organisms containing glycoproteins with n-linked glycans.

25 The term "plant" as used herein refers to a plurality of plant cells which are largely differentiated into a structure that is present at any stage of a plant's development. Such structures include, but are not limited to, a fruit, shoot, stem, root, leaf, seed, flower petal, or similar structure. The term "plant tissue" includes differentiated and undifferentiated tissues of plants including, but not limited to, roots, shoots, leaves, pollen, seeds, tumor tissue and various types of cells in culture 30 (e.g., single cells, protoplasts, embryos, callus, protocorm-like bodies, and other types of cells). Plant tissue may be *in planta*, in organ culture, tissue culture, or cell culture. Similarly, "plant cells" may be cells in culture or may be part of a plant.

Glycosyltransferases are enzymes that catalyze the processing reactions that determine the 35 structures of cellular oligosaccharides, including the oligosaccharides on glycoproteins. As used herein, "glycosyltransferase" is meant to include mannosidases, even though these enzymes trim glycans and do not "transfer" a monosaccharide. Glycosyltransferases share the feature of a type II membrane orientation. Each glycosyltransferase is comprised of an amino terminal cytoplasmic tail (shown for illustration purposes below as a made up of a string of amino acids arbitrarily labeled "X"

– without intending to suggest the actual size of the region), a signal anchor domain (shown below as made up of a string of amino acids labeled “H” for hydrophobic – without intending to suggest the actual size of the domain and without intending to suggest that the domain is only made up of hydrophobic amino acids) that spans the membrane (referred to herein as a “transmembrane domain”), followed by a luminal stem (shown below as made up of a string of amino acids arbitrarily labeled “S” – without intended to suggest the actual size of the region) or stalk region, and a carboxy-terminal catalytic domain (shown below as made up of a string of amino acids arbitrarily labeled “C” – without intending to suggest the actual size of the domain):



10 Collectively, The Cytoplasmic Tail-Transmembrane-Stem Region or “CTS” (which has been underlined in the above schematic for clarity) can be used (or portions thereof) in embodiments contemplated by the present invention wherein the catalytic domain is exchanged or “swapped” with a corresponding catalytic domain from another molecule (or portions of such regions/domains) to create a hybrid protein.

15 For example, in a preferred embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme (as well as vectors containing such nucleic acid, host cells containing such vectors, and the hybrid enzyme itself), said hybrid enzyme comprising at least a portion of a CTS region [*e.g.*, the cytoplasmic tail (“C”), the transmembrane domain (“T”), the cytoplasmic tail together with the transmembrane domain (“CT”), the transmembrane domain together with the stem (“TS”), or the complete CTS region] of a first glycosyltransferase (*e.g.* plant glycosyltransferase) and at least a portion of a catalytic region of a second glycosyltransferase (*e.g.* mammalian glycosyltransferase). To create such an embodiment, the coding sequence for the entire CTS region (or portion thereof) may be deleted from nucleic acid coding for the mammalian glycosyltransferase and replaced with the coding sequence for the entire CTS region (or portion thereof) of a plant glycosyltransferase. On the other hand, a different approach might be taken to create this embodiment; for example, the coding sequence for the entire catalytic domain (or portion thereof) may be deleted from the coding sequence for the plant glycosyltransferase and replaced with the coding sequence for the entire catalytic domain (or portion thereof) of the mammalian glycosyltransferase. In such a case, the resulting hybrid enzyme would have the amino-terminal cytoplasmic tail of the plant glycosyltransferase linked to the plant glycosyltransferase transmembrane domain linked to the stem region of the plant glycosyltransferase in the normal manner of the wild-type plant enzyme – but the stem region would be linked to the catalytic domain of the mammalian glycosyltransferase (or portion thereof).

30 35 It is not intended that the present invention be limited only to the two approaches outlined above. Other variations in the approach are contemplated. For example, to create nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian

glycosyltransferase, one might use less than the entire coding sequence for the CTS region (*e.g.*, only the transmembrane domain of the plant glycosytransferase, or the complete cytoplasmic tail together with all or a portion of the transmembrane domain, or the complete cytoplasmic tail together with all of the transmembrane domain together with a portion of the stem region). One might delete the
5 mammalian coding sequence for the entire cytoplasmic tail together with the coding sequence for the transmembrane domain (or portion thereof) – followed by replacement with the corresponding coding sequence for the cytoplasmic tail and transmembrane domain (or portion thereof) of the plant glycosyltransferase. In such a case, the resulting hybrid enzyme would have the stem region of the mammalian glycosyltransferase linked to the plant glycosyltransferase transmembrane domain (or portion thereof) which in turn would be linked to the amino-terminal cytoplasmic tail of the plant
10 glycosyltransferase, with the stem region being linked to the catalytic domain of the mammalian glycosyltransferase (*i.e.* two of the four regions/domains would be of plant origin and two would be of mammalian origin).

In other embodiments, the present invention contemplates nucleic acid encoding a hybrid
15 enzyme (along with vectors, host cells containing the vectors, plants – or plant parts - containing the host cells), said hybrid enzyme comprising at least a portion of an amino-terminal cytoplasmic tail of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase. In this embodiment, the hybrid enzyme encoded by the nucleic acid might or might not contain other plant sequences (*e.g.*, the transmembrane domain or portion thereof, the stem
20 region or portion thereof). For example, to create such an embodiment, the coding sequence for the entire cytoplasmic tail (or portion thereof) may be deleted from nucleic acid coding for the mammalian glycosyltransferase and replaced with the coding sequence for the entire cytoplasmic domain (or portion thereof) of a plant glycosyltransferase. In such a case, the resulting hybrid enzyme would have the amino-terminal cytoplasmic tail (or portion thereof) of the plant
25 glycosyltransferase linked to the mammalian glycosyltransferase transmembrane domain, which in turn is linked to stem region of the mammalian glycosyltransferase, the stem region being linked to the catalytic domain of the mammalian glycosyltransferase. On the other hand, a different approach might be taken to create this embodiment; for example, the coding sequence for the entire catalytic domain (or portion thereof) may be deleted from the coding sequence for the plant
30 glycosyltransferase and replaced with the coding sequence for the entire catalytic domain (or portion thereof) of the mammalian glycosyltransferase. In such a case, the resulting hybrid enzyme would have the amino-terminal cytoplasmic tail of the plant glycosyltransferase linked to the plant glycosyltransferase transmembrane domain linked to the stem region of the plant glycosyltransferase in the normal manner of the wild-type plant enzyme – but the stem region would be linked to the
35 catalytic domain of the mammalian glycosyltransferase (or portion thereof).

In the above discussion, the use of the phrase “or portion thereof” was used to expressly indicate that less than the entire region/domain might be employed in the particular case (*e.g.*, a fragment might be used). For example, the cytoplasmic tail of glycosyltransferases ranges from

approximately 5 to 50 amino acids in length, and more typically 15 to 30 amino acids, depending on the particular transferase. A "portion" of the cytoplasmic tail region is herein defined as no fewer than four amino acids and can be as large as up to the full length of the region/domain less one amino acid. It is desired that the portion function in a manner analogous to the full length region/domain – but need not function to the same degree. For example, to the extent the full-length cytoplasmic tail functions as a Golgi retention region or ER retention signal, it is desired that the portion employed in the above-named embodiments also function as a Golgi or ER retention region, albeit perhaps not as efficiently as the full-length region.

Similarly, the transmembrane domain is typically 15-25 amino acids in length and made up of primarily hydrophobic amino acids. A "portion" of the transmembrane domain is herein defined as no fewer than ten amino acids and can be as large as up to the full length of the region/domain (for the particular type of transferase) less one amino acid. It is desired that the portion function in a manner analogous to the full length region/domain – but need not function to the same degree. For example, to the extent the full-length transmembrane domain functions as the primary Golgi retention region or ER retention signal, it is desired that the portion employed in the above-named embodiments also function as a Golgi or ER retention region, albeit perhaps not as efficiently as the full-length region. The present invention specifically contemplates conservative substitutions to create variants of the wild-type transmembrane domain or portions thereof. For example, the present invention contemplates replacing one or more hydrophobic amino acids (shown as "H" in the schematic above) of the wild-type sequence with one or more different amino acids, preferably also hydrophobic amino acids.

A portion of the catalytic domain can be as large as the full length of the domain less one amino acid. Where the catalytic domain is from a beta1,4-galactosyltransferase, it is preferred that the portion include at a minimum residues 345-365 which are believed to be involved in the conformation conferring an oligosaccharide acceptor binding site (it is preferred that the portion include this region at a minimum and five to ten amino acids on either side to permit the proper conformation).

The present invention also includes synthetic CTS regions and portions thereof. A "portion" of a CTS region must include at least one (and may include more than one) entire domain (e.g., the entire transmembrane domain) but less than the entire CTS region.

Importantly, by using the term "CTS region" or "transmembrane domain" it is not intended that only wild type sequences be encompassed. Indeed, this invention is not limited to natural glycosyltransferases and enzymes involved in glycosylation, but also includes the use of synthetic enzymes exhibit the same or similar function. In one embodiment, wild type domains are changed (e.g. by deletion, insertion, replacement and the like).

Finally, by using the indicator "Tm" when referring to a particular hybrid (e.g., "TmXyl-), entire transmembrane/CTS domains (with or without changes to the wild-type sequence) as well as portions (with or without changes to the wild-type sequence) are intended to be encompassed.

SUMMARY OF THE INVENTION

The present invention contemplates nucleic acid (whether DNA or RNA) encoding hybrid enzymes (or "fusion proteins"), vectors containing such nucleic acid, host cells (including but not limited to cells in plant tissue and whole plants) containing such vectors and expressing the hybrid enzymes, and the isolated hybrid enzyme(s) themselves. In one embodiment, expression of said hybrid enzymes (or "fusion proteins") results in changes in glycosylation, such as, but not limited to, reduction of sugar moieties such as xylose, fucose, Lewis^{A/B/X} or other sugar structures that interfere with desired glycoform accumulation. In one embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region (or portion thereof) of a glycosyltransferase (including but not limited to a plant glycosyltransferase) and a catalytic region (or portion thereof) of a non-plant glycosyltransferase (e.g., mammalian, fish, amphibian, fungal). It is preferred that, when expressed, the CTS region (or portion thereof) is linked (directly or indirectly) in operable combination to said catalytic region (or portion thereof).

10 The linking is preferably covalent and the combination is operable in that the catalytic region exhibits catalytic function (even if said catalytic function is reduced as compared to the wild-type enzyme). The linking can be direct in the sense that there are no intervening amino acids or other regions/domains. On the other hand, the linking can be indirect in that there are intervening amino acids (or other chemical groups) and/or other regions/domains between them. Of course, the nucleic acid used to make the nucleic acid encoding the above-described hybrid enzyme(s) can be obtained enzymatically from a physical sequence (e.g. genomic DNA, a cDNA, and the like) or alternatively, made synthetically using a reference sequence (e.g. electronic or hardcopy sequence) as a guide.

15 20 25

In a particular embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region (e.g., at least a transmembrane region and optionally more of the CTS region) of a plant glycosyltransferase and a catalytic region (or portion thereof) of a non-plant (such as a mammalian) glycosyltransferase. Again, it is preferred that, when expressed, these regions are linked (directly or indirectly) in operable combination. In yet another embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane domain (or portion thereof) of a plant glycosyltransferase and a catalytic region (or portion thereof) of a mammalian glycosyltransferase. Again, it is preferred that, when expressed, these regions are linked (directly or indirectly) in operable combination.

30 35 It is not intended that the present invention be limited to particular transferases. In one embodiment, the plant glycosyltransferase is a xylosyltransferase. In another embodiment, the plant glycosyltransferase is a N-acetylglucosaminyltransferase. In another embodiment, the plant glycosyltransferase is a fucosyltransferase. In a preferred embodiment, the mammalian glycosyltransferase is a human galactosyltransferase (such as the human beta 1,4-

galactosyltransferase encoded by SEQ ID NO:1 wherein the nucleotides encoding the transmembrane domain are deleted and replaced).

It is not intended that the present invention is limited to the use of a plant-derived glycosyltransferase CTS-domain and a human glycosyltransferase catalytic domain but also vice versa and the use of any CTS-domain of a glycosyltransferase in combination with the catalytic fragment of at least one other glycosyltransferase. Indeed, the present invention broadly contemplates, in one embodiment, nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase. It is preferred that said first and second glycosyltransferases are from different species (and can be from a different genus or even from a different phylum). In one embodiment, said first glycosyltransferase comprises a plant glycosyltransferase. In another embodiment, said plant glycosyltransferase is a xylosyltransferase. In yet another embodiment, said plant glycosyltransferase is a fucosyltransferase. In a preferred embodiment said second glycosyltransferase comprises a mammalian glycosyltransferase. In a particularly preferred embodiment, said mammalian glycosyltransferase is a human galactosyltransferase.

It is not intended that the present invention be limited to circumstances where the first and second glycosyltransferases are plant and non-plant, respectively. In one embodiment, said first glycosyltransferase comprises a first mammalian glycosyltransferase and said second glycosyltransferase comprises a second mammalian glycosyltransferase. In a preferred embodiment, said first mammalian glycosyltransferase is a non-human glycosyltransferase and said second mammalian glycosyltransferase is a human glycosyltransferase.

It is not intended that the present invention be limited to the type of vector. In one embodiment, the present invention contemplates an expression vector, comprising the nucleic acid encoding the above-described hybrid enzyme.

It is also not intended that the present invention be limited to the type of host cells. A variety of prokaryotic and eukaryotic host cells are commercially available for expressing proteins. In one embodiment, the present invention contemplates a host cell containing the vector comprising the nucleic acid encoding the above-described hybrid enzyme (with or without other vectors or other nucleic acid encoding other hybrid enzymes or glycosyltransferases). In a preferred embodiment, the host cell is a plant cell. In a particularly preferred embodiment, the present invention contemplates a plant comprising such a host cell.

It is not intended that the present invention be limited by the method by which host cells are made to express the hybrid enzymes of the present invention. In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a host cell (such as a plant cell, whether in culture or as part of plant tissue or even as part of an intact growing plant), and ii) an expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a CTS region of a plant glycosyltransferase (e.g. the transmembrane domain) and at least a portion of a catalytic region of a mammalian glycosyltransferase; and b)

introducing said expression vector into said plant cell under conditions such that said hybrid enzyme is expressed. Again, it is not intended that the present invention be limited to particular transferases. In one embodiment, the plant glycosyltransferase used in the above-described method is a xylosyltransferase. In another embodiment, the plant glycosyltransferase is a N-
5 acetylglucosaminyltransferase. In another embodiment, the plant glycosyltransferase is a fucosyltransferase. In a preferred embodiment, the mammalian glycosyltransferase used in the above-described method is a human galactosyltransferase (such as the human beta 1,4-galactosyltransferase encoded by SEQ ID NO:1 wherein the nucleotides encoding the transmembrane domain are deleted and replaced) (or simply where the nucleotides of SEQ ID NO:1
10 encoding the catalytic domain, or portion thereof, are taken and linked to nucleotides encoding the CTS region, or portion thereof, of a plant glycosyltransferase.).

It is not intended that the present invention be limited to a particular scheme for controlling glycosylation of a heterologous protein using the hybrid enzymes described above. In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a host cell (such as a plant cell), ii) a first expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a CTS region (e.g. at least a transmembrane domain) of a first (such as a plant) glycosyltransferase and at least a portion of a catalytic region of a second (such as a mammalian) glycosyltransferase, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; (or portion thereof; and b) introducing said first and second expression vectors into said plant cell under conditions such that said hybrid enzyme and said heterologous protein are expressed. Alternatively, a single vector with nucleic acid encoding both the hybrid enzyme (or hybrid enzymes) and the heterologous glycoprotein might be used. Regardless of which method is used, the invention contemplates, in one embodiment, the additional step (c) of isolating the heterologous protein – as well as the isolated protein itself as a composition.
15
20
25

On the other hand, the present invention also contemplates introducing different vectors into different plant cells (whether they are cells in culture, part of plant tissue, or even part of an intact growing plant). In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a first plant comprising a first expression vector, said first vector comprising nucleic acid encoding a hybrid enzyme (or encoding two or more hybrid enzymes), said hybrid enzyme comprising at least a portion of a CTS region (e.g. the first approximately 40-60 amino acids of the N-terminus) of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, and ii) a second plant comprising a second expression vector, said second vector comprising nucleic acid encoding a heterologous protein (or portion thereof); and crossing said first plant and said second plant to produce progeny expressing said hybrid enzyme and said heterologous protein. Of course, such progeny can be isolated, grown up, and analyzed for the presence of each (or both) of the proteins. Indeed, the heterologous protein can be used (typically first purified substantially free of plant cellular material) therapeutically (e.g., administered to a human or animal,
30
35

whether orally, by intravenous, transdermally or by some other route of administration) to treat or prevent disease.

It is not intended that the present invention be limited to a particular heterologous protein. In one embodiment, any peptide or protein that is not endogenous to the host cell (or organism) is contemplated. In one embodiment, the heterologous protein is an antibody or antibody fragment. In a particularly preferred embodiment, the antibody is a human antibody or "humanized" antibody expressed in a plant in high yield. "Humanized" antibodies are typically prepared from non-human antibodies (*e.g.* rodent antibodies) by taking the hypervariable regions (the so-called CDRs) of the non-human antibodies and "grafting" them on to human frameworks. The entire process can be synthetic (provided that the sequences are known) and frameworks can be selected from a database of common human frameworks. Many times, there is a loss of affinity in the process unless either the framework sequences are modified or the CDRs are modified. Indeed, increases in affinity can be revealed when the CDRs are systematically mutated (for example, by randomization procedures) and tested.

While the present invention is particularly useful in the context of heterologous proteins, in one embodiment, the hybrid enzymes of the present invention are used to change the glycosylation of endogenous proteins, *i.e.* proteins normally expressed by the host cell or organism.

The present invention specifically contemplates the plants themselves. In one embodiment, the present invention contemplates a plant, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a CTS region (*e.g.* the cytoplasmic tail together with at least a portion of the transmembrane domain) of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, said second expression vector, said second vector comprising nucleic acid encoding a heterologous protein (or portion thereof). In a preferred embodiment, by virtue of being expressed along with the hybrid enzyme (or hybrid enzymes) of the present invention, the heterologous protein displays reduced (10% to 99%) alpha 1,3 -fucosylation (or even no fucosylation), as compared to when the heterologous protein is expressed in the plant in the absence of the hybrid enzyme (or enzymes). In a preferred embodiment, by virtue of being expressed along with the hybrid enzyme (or hybrid enzymes) of the present invention, the heterologous protein displays reduced fucose and xylose, as compared to when the heterologous protein is expressed in the plant in the absence of the hybrid enzyme (or enzymes).

It is not intended that the present invention be limited to a particular theory by which reduced fucose and/or xylose is achieved. Very little is known about the sub-Golgi sorting mechanism in plants. The mammalian specific $\beta(1,4)$ -galactosyltransferase (GalT) has been used (see the

Examples below) as an excellent first marker to study this phenomenon since it generates glycan structures not normally found in plants. The glycan structures of plants that express galactosyltransferase has been compared with glycan structures from plants that express a chimeric galactosyltransferase of which the CTS domain is exchanged for that of a plant xylosyltransferase (or portion thereof). The change in observed glycan structures show that the galactosyltransferase is, as in mammals, confined to a specific sub-compartment of the plant Golgi. Without limiting the invention to any particular mechanism, the sorting mechanism of plants and mammals are apparently conserved even to the extent that glycosyltransferases unknown to plants are routed to specific analogous location in the Golgi. This location is later in the Golgi than where the endogenous 5 xylosyl-, fucosyl- and GlcNAcTII (GnTII) transferases are located.

10 The finding that N-glycans in these plants that express relocalised variants of GalT containing significantly less xylose and fucose is also of biotechnological relevance. For 15 glycoproteins intended for therapeutic use in mammals, such as humans, the approach of certain embodiments of the present invention provides methods and compositions for controlling N-linked glycosylation of glycoproteins in plants so that glycoprotein essentially free of xylose and fucose and containing at least a bi-antennary N-glycans (but not limited to bi-antennary, also include tri-antennary, and the like) and (at least one) galactose residue on at least one of the arms of the N-glycan can be obtained. Hence, it is not intended that the present invention is limited to bi-antennary 20 N-glycans but also includes bisected bi-antennary N-glycans, tri-antennary N-glycans, and the like. Furthermore, the invention is not limited to complex-type N-glycans but also includes hybrid-type N-glycans and other type N-glycans. The present invention contemplates such resulting glycoproteins. In addition, the methods and compositions of the present invention may be applicable for plants and non-plant systems where besides xylose, fucose, Lewis^{A/B/X} type N-glycan modifications (β1-3-GalT, α1-4-FucT, other) or other sugars, "interfere" with desired glycoform accumulation.

25 In one embodiment, the invention is directed to controlling N-linked glycosylation of plants by modulating the localization of enzymes involved in glycan biosynthesis in the Golgi apparatus. Specifically, embodiments of the invention are directed to a method of producing in a plant host system a glycoprotein having bi-antennary glycans and containing at least one galactose residues on at least one of the arms and which are devoid (or reduced in) of xylose and fucose, comprising:(a) preventing (or inhibiting) addition of xylose and fucose on the core of the glycan of said glycoprotein 30 and (b) adding one or preferably two galactose residues to said arms.

35 Addition of xylose and fucose to said heterologous glycoprotein may be reduced or even prevented by introducing to said plant host system a nucleic acid encoding a hybrid enzyme comprising a CTS region (or portion thereof) of a protein, particularly an enzyme such as plant xylosyltransferase and catalytic region (or portion thereof) of a galactosyltransferase not normally found in a plant, or a modified galactosyltransferase where its transmembrane portion has been removed and endoplasmic reticulum retention signal have been inserted, wherein said protein or enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said

galactosyltransferase. It is preferred that the galactosyltransferase is a mammalian galactosyltransferase and in particular, a human galactosyltransferase. In a most specific embodiment, said galactosyltransferase is human β 1,4 galactosyltransferase (GalT). In a preferred embodiment, said xylosyltransferase is a β 1,2-xylosyltransferase. The exchange of the CTS region or 5 CTS fragment of a mammalian glycosyltransferase (such as a galactosyltransferase) by one from the group of enzymes that act earlier in the Golgi apparatus than galactosyltransferase including but not limited to those from of XylT, FucT, GnTI, GnTII, GnTIII, GnTIV, GnTV, GnTVI, ManI, ManII and ManIII results in strongly reduced amounts of glycans that contain the undesired xylose and fucose residues (see Figure 2). In addition, galactosylation is improved and the diversity in glycans is 10 reduced. While not limited to any particular mechanism, the increase in galactosylated glycans that carry neither xylose nor fucose is believed to be mainly attributed to the accumulation of GalGNMan5, GNMan5 or GalGNMan4. Also, galactosylation occurs on one glycan arm only. Apparently, the galactosylation earlier in the Golgi inhibits trimming of the said glycoforms by Mannosidase II (ManII) to GalGNMan3. Also addition of the second GlcNAc by GlcNAcTII 15 (GnTII) is inhibited.

Therefore, in one embodiment, a further step is contemplated to obtain the desired glycoprotein that has both arms galactosylated and yet is essentially devoid of xylose and fucose. Thus, in one embodiment, the method of the invention as noted above further comprises adding galactose residues to the arms of said glycoprotein (see Figure 3). In one embodiment of the 20 invention, galactose residues are added onto both arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of GnTI and the active domain (or portion thereof) of GnTII; (b) a nucleic acid sequence encoding the second hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane of GnTI and the active domain of ManII and 25 (c) a nucleic acid sequence encoding a third hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of XylT and the active domain (or portion thereof) of human galactosyltransferase (TmXyl-GalT). In another embodiment of the invention, galactose residues are added onto both arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising the CTS region (or fragment, such as one 30 including the transmembrane domain) of ManI and the active domain (or portion thereof) of GnTI; (b) a nucleic acid sequence encoding the second hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of ManI and the active domain (or portion thereof) of GnTII; (c) a nucleic acid sequence encoding the third hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of ManI and the 35 active domain (or portion thereof) of ManII, and (d) a nucleic acid sequence encoding a fourth hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of XylT and the active domain (or portion thereof) of human galactosyltransferase (TmXyl-GalT).

It is not intended that the present invention be limited to particular combinations of hybrid enzymes or the number of such hybrid enzymes employed in a single cell, plant tissue or plant. In a preferred embodiment, the present invention contemplates host cells expressing TmXyl-GalT plus TmGnTI-GnTII plus TmGnTI-ManII. In one embodiment of the invention, galactose residues are added to said arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising a CTS region (or fragment thereof) of a protein, particularly an enzyme, including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region (or portion thereof) of a mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said mannosidase II or modified mannosidase II where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted and (b) a nucleic acid sequence encoding a second hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of an enzyme including but not limited to N-acetyl-glucosaminyltransferase I (GnTI) and a catalytic region (or portion thereof) of a N-acetylglucosaminyl-transferase II (GnTII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N acetylglucosaminyl-transferaseII (GnTII) or modified N-acetylglucosaminyltransferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted. The sequences encoding N-acetylglucosaminyltransferases or mannosidase II or the said transmembrane fragments can originate from plants or from eukaryotic non-plant organisms (e.g., mammals).

In yet another preferred embodiment, the present invention contemplates a host cell expressing TmXyl-GalT plus TmManI-GnTI plus TmManI-ManII plus TmManI-GnTII. In another embodiment of the invention, galactose residues are added to said arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of a protein, particularly an enzyme, including but not limited to Mannosidase I (ManI) and a catalytic region (or portion thereof) of a N acetylglucosaminyltransferase I (GnTI), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N-acetylglucosaminyl-transferase I (GnTI) or modified N acetylglucosaminyltransferase I (GnTI) where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted and (b) a nucleic acid sequence encoding a second hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of an enzyme including but not limited to Mannosidase I (ManI) and a catalytic region (or portion thereof) of a Mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said Mannosidase II (ManII) or modified Mannosidase II (ManII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted and (c) a nucleic acid sequence encoding a third hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of an enzyme including but not limited to Mannosidase I (ManI) and a catalytic region (or portion thereof) of a N-acetylglucos-aminyltransferase II (GnTII), wherein said

enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N-acetylglucosaminyltransferase II (GnTII) or modified N-acetylglucosaminyltransferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted. The sequences encoding N-acetylglucosaminyltransferases or mannosidases or 5 the said transmembrane fragments can originate from plants or from eukaryotic non-plant organisms (e.g., mammals).

In still another preferred embodiment, the present invention contemplates host cells expressing TmXyl-GalT plus ManIII. In another embodiment of the invention, galactose residues are added to said arms by introducing to said plant host system (a) a nucleic acid sequence encoding 10 a Mannosidase III (ManIII, wildtype gene sequence but not limited to: also ManIII with endoplasmic reticulum retention signal; ManIII with transmembrane fragment of early (cis-) Golgi apparatus glycosyltransferase (GnTI, ManI, GnTIII). The sequences encoding Mannosidase III can originate form insects, preferably from *Spodoptera frugiperda* or *Drosophila melanogaster* (but not limited to), human or from other organisms.

15 In still another preferred embodiment, the present invention contemplates a host cell expressing TmXyl-GalT plus ManIII plus TmGnTI-GnTII. In yet another preferred embodiment, the present invention contemplates a host cell expressing TmXyl-GalT plus ManIII plus TmManI-GnTI plus TmManI-GnTII.

The method of the invention may optionally comprise, in one embodiment, introducing into 20 said plant host system a mammalian N-acetylglucosaminyltransferase GnTIII, particularly a human GnTIII or hybrid protein comprising a catalytic portion of mammalian GnTIII and a transmembrane portion of a protein, said protein residing in the ER or earlier compartment of the Golgi apparatus of a eukaryotic cell. For example, in one embodiment, the hybrid enzyme TmXyl-GnTIII is contemplated (along with nucleic acid coding for such a hybrid enzyme, vectors containing such 25 nucleic acid, host cells containing such vectors, and plants – or plant parts – containing such host cells). In another embodiment, the hybrid enzyme TmFuc-GnTIII is contemplated (along with nucleic acid coding for such a hybrid enzyme, vectors containing such nucleic acid, host cells containing such vectors, and plants – or plant parts – containing such host cells). The present invention specifically contemplates host cells expressing such hybrid enzymes (with or without 30 additional hybrid enzymes or other glycosyltransferases).

The invention is further directed to said hybrid and modified enzymes, nucleic acid sequences encoding said hybrid enzymes, vectors comprising said nucleic acid sequences and methods for obtaining said hybrid enzymes. Furthermore, the invention is directed to a plant host system comprising a heterologous glycoprotein having preferably complex type bi-antennary glycans and containing at least one galactose residue on at least one of the arms and are devoid of xylose and 35 fucose. A “heterologous glycoprotein” is a glycoprotein originating from a species other than the

plant host system. The glycoprotein may include but is not limited to antibodies, hormones, growth factors and growth factor receptors and antigens.

Indeed, the present invention is particularly useful for controlling the glycosylation of heterologous glycoproteins, such as antibodies or antibody fragments (single chain antibodies, Fab 5 fragments, Fab₂ fragments, Fv fragments, and the like). To control the glycosylation of an antibody, the gene construct encoding a hybrid enzyme of the present invention (e.g., the TmXyl-GalT gene construct) can be introduced in transgenic plants expressing an antibody (e.g., monoclonal antibody) or antibody fragment. On the other hand, the gene(s) encoding the antibody (or antibody fragment) can be introduced by retransformation of plant expressing TmXyl-GalT gene construct. In still 10 another embodiment, the binary vector harbouring the TmXyl-GalT expression cassette can be co-transformed to plants together with a plant binary vector harbouring the expression cassettes comprising both light and heavy chain sequences of a monoclonal antibody on a single T-DNA or with binary vectors harbouring the expression cassettes for light and heavy chain sequences both separately on independent T-DNA's but both encoding a monoclonal antibody. The present 15 invention specifically contemplates, in one embodiment, crossing plants expressing antibodies with plant expressing the hybrid glycosyltransferase(s) of the present invention.

A "host system" may include but is not limited to any organism containing glycoproteins with N-glycans.

A "plant host system" may include but is not limited to a plant or portion thereof, which 20 includes but is not limited to a plant cell, plant organ and/or plant tissue. The plant may be a monocotyledon (monocot) which is a flowering plant whose embryos have one cotyledon or seed leaf and includes but is not limited to lilies, grasses, corn (*Zea mays*), rice, grains including oats, wheat and barley, orchids, irises, onions and palms. Alternatively, the plant may be a dicotyledon (dicot) which includes but is not limited to tobacco (*Nicotiana*), tomatoes, potatoes, legumes (e.g., 25 alfalfa and soybeans), roses, daisies, cacti, violets and duckweed. The plant may also be a moss which includes but is not limited to *Physcomitrella patens*.

The invention is further directed to a method for obtaining said plant host system. The method comprises crossing a plant expressing a heterologous glycoprotein with a 30 plant comprising (a) a hybrid enzyme comprising a catalytic region (or portion thereof) of a galactosyltransferase not normally found in a plant and a CTS region (or fragment, such as one including the transmembrane domain) of a protein, wherein said protein acts earlier in the Golgi apparatus of a plant cell in said plant host system than said galactosyltransferase or a modified galactosyltransferase where its transmembrane portion has been deleted and endoplasmic reticulum retention signal has been inserted; (b) a hybrid enzyme comprising a CTS region (or portion thereof, 35 such as one including the transmembrane domain) of a protein, particularly an enzyme, including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region (or portion thereof) of a mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said mannosidase II or modified mannosidase II where its transmembrane

portion has been deleted and endoplasmic reticulum retention signal have been inserted and (c) a hybrid enzyme comprising at least a transmembrane region of an enzyme (such as the first 40-60 amino acids of the N-terminus) of a glycosyltransferase including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region of a N-acetylglucosaminyltransferase

5 II (GnTII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N acetylglucosaminyltransferase II (GnTII) or modified N-acetylglucosaminyl-transferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted., harvesting progeny from said crossing and selecting a desired progeny plant expressing said heterologous glycoprotein.

10 The invention is further directed to said plant or portion thereof which would constitute a plant host system. Said plant host system may further comprise a mammalian GnTIII enzyme or hybrid protein comprising a catalytic portion of mammalian GnTIII and a transmembrane portion of a protein, said protein residing in the ER or earlier compartment of the Golgi apparatus of a eukaryotic cell.

15 Additionally, the invention also provides the use of a plant host system to produce a desired glycoprotein or functional fragment thereof. The invention additionally provides a method for obtaining a desired glycoprotein or functional fragment thereof comprising cultivating a plant according to the invention until said plant has reached a harvestable stage, for example when sufficient biomass has grown to allow profitable harvesting, followed by harvesting said plant with
20 established techniques known in the art and fractionating said plant with established techniques known in the art to obtain fractionated plant material and at least partly isolating said glycoprotein from said fractionated plant material.

Alternatively, said plant host cell system comprising said heterologous glycoprotein may also be obtained by introducing into a plant host cell system or portion thereof (a) a nucleic acid sequence encoding a hybrid enzyme comprising a catalytic region of a galactosyltransferase not normally found in a plant and at least the transmembrane region (or more of the CTS) of a protein, wherein said protein acts earlier in the Golgi apparatus of a plant cell in said plant host system than said galactosyltransferase or a modified galactosyltransferase where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted; (b) a nucleic acid sequence encoding a first hybrid enzyme comprising at least the transmembrane region (or more of the CTS if desired) of a protein, particularly an enzyme, including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region of a mannosidase II (ManII) , wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said mannosidase II, or modified mannosidase II where its transmembrane portion has been deleted and
30 endoplasmic reticulum retention signal have been inserted and (c) a nucleic acid sequence encoding a second hybrid enzyme comprising at least a transmembrane region (more of the CTS if desired) of an enzyme including but not limited to N-acetylglucosaminyl-transferase I (GnTI) and a catalytic region of a N-acetylglucosaminyltransferase II (GnTII), wherein said enzyme acts earlier in the

Golgi apparatus of a plant cell in said plant host system than said N- acetylglucosaminyltransferase-II (GnTII) or modified N-acetylglucosaminyltransferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted. and isolating a plant or portion thereof expressing said heterologous glycoprotein (or portion thereof). In one embodiment, one vector comprising all of the nucleic acid sequences is introduced into said plant host system. In another embodiment, each nucleic acid sequence is inserted into separate vectors and these vectors are introduced into said plant host system. In another embodiment combinations of two or more nucleic acid sequences are inserted into separate vectors which are then combined into said plant host system by retransformation or co-transformation or by crossing.

The invention also provides use of such a plant-derived glycoprotein or functional fragment thereof according to the invention for the production of a composition, particularly, pharmaceutical composition, for example for the treatment of a patient with an antibody, a hormone, a vaccine antigen, an enzyme, or the like. Such a pharmaceutical composition comprising a glycoprotein or functional fragment thereof is now also provided.

Finally, it is contemplated that the above-described approach may be useful in reducing the overall diversity in glycans in plants expressing one or more of the hybrid enzymes of the present invention (as compared to wild-type plants or plants simply transformed with only mammalian GalT).

20 BRIEF DESCRIPTION OF THE FIGURES

Figure 1 compares the glycosylation pathway of glycoproteins in plants and in mammals.

Figure 2 shows the effect of exchanging the CTS fragment of galactosyltransferase with xylosyltransferase

Figure 3 shows the further effect of relocating mannosidase II and GlcNAcTII.

Figure 4 top panel shows a T-DNA construct carrying the genes encoding glycan modifying enzymes to produce efficiently galactosylated glycans that are devoid of immunogenic xylose and fucose and the bottom panel shows a T-DNA construct carrying antibody light chain and heavy chain genes.

Figure 5 shows the nucleic acid sequence (SEQ ID NO:1) for a human galactosyltransferase (human B1,4-galactosyltransferase – GalT).

Figure 6 shows the nucleic acid sequence of Figure 5 along with the corresponding amino acid sequence (SEQ ID NO:2).

Figure 7 shows an illustrative mutated sequence (SEQ ID NO:59) derived the wild type amino acid sequence (SEQ ID NO:2) for a human galactosyltransferase, wherein a serine has been deleted from the cytoplasmic tail and a G-I-Y motif has been repeated. Of course, such changes are merely illustrative of the many possible changes within the scope of the present invention. For example, in one embodiment, the present invention contemplates mutated sequences wherein only deletions (one or more) are employed (e.g. deletions in the cytoplasmic tail domain or the stem

domain) – with no insertions or repeats. Similarly, in one embodiment, the present invention contemplates mutated sequences wherein only (one or more) insertions or replacements (e.g. in the transmembrane domain) are employed – with no deletions.

Figure 8 shows the nucleic acid sequence (SEQ ID NO:3) encoding a hybrid enzyme comprising human galactosyltransferase (human B1,4-galactosyltransferase – GalT). The upper case letters are nucleotides of *Arabidopsis thaliana* mRNA for beta 1,2-xylosyltransferase (database entry: EMBL:ATH277603, the TmXyl-fragment used involves nucleotides 135-297 of this database sequence).

Figure 9 shows the nucleic acid sequence of Figure 8 along with the corresponding amino acid sequence (SEQ ID NO:4).

Figure 10 shows the amino acid sequence (SEQ ID NO:4) for the hybrid enzyme encoded by the nucleic acid shown in Figure 8.

Figure 11 shows the nucleic acid sequence (SEQ ID NO:5) for the human glycosyltransferase GnTIII (along with additional sequence encoding a myc-tag) (primary accession number Q09327 GNT3 HUMAN).

Figure 12 shows the nucleic acid sequence of Figure 11 along with the corresponding amino acid sequence (SEQ ID NO:6).

Figure 13 shows the amino acid sequence (SEQ ID NO:6) for a human GnTIII (along with additional amino acid sequence of the myc epitope tag SEQ ID NO:7).

Figure 14 shows the nucleic acid sequence (SEQ ID NO:9) encoding one embodiment of a hybrid enzyme of the present invention, said hybrid enzyme comprising the transmembrane domain of a plant xylosyltransferase (TmXyl-) and the catalytic domain (along with other regions) for human GnTIII (TmXyl-GnTIII) (along with additional sequence encoding a myc-tag).

Figure 15 shows the nucleic acid sequence of Figure 14 along with the corresponding amino acid sequence (SEQ ID NO:10).

Figure 16 shows the amino acid sequence (SEQ ID NO:10) for hybrid enzyme encoded by the nucleic acid of Figure 14 (along with additional sequence for the myc epitope tag SEQ ID NO:7).

Figure 17 shows the complete nucleic acid sequence (SEQ ID NO:27) for a cassette encoding the hybrid enzymes TmXyl-GaIT plus TmGnTI-GnTII plus TmGnTI-ManII).

Figure 18 shows the complete nucleic acid sequence (SEQ ID NO:28) for a cassette encoding the hybrid enzyme TmGnTI-ManII (with the RbcS1 promoter sequence SEQ ID NO:39 shown).

Figure 19 shows the nucleic acid sequence (SEQ ID NO:29) encoding the hybrid enzyme TmGnTI-ManII.

Figure 20 shows the nucleic acid sequence (SEQ ID NO:30) encoding the hybrid enzyme TmGnTI-GnTII.

Figure 21 shows the nucleic acid sequence (SEQ ID NO:31) encoding the hybrid enzyme TmGnTI-GnTII, wherein the transmembrane fragment used (designated TmGntI) has the nucleic acid sequence set forth in SEQ ID NO:32.

Figure 22A shows the nucleic acid sequence (SEQ ID NO:32) encoding one embodiment of a transmembrane domain fragment (TmGnTI). Figure 22B shows the nucleic acid sequence (SEQ ID NO:33) encoding another embodiment of a transmembrane domain fragment (TmManI).

Figure 23 shows the complete nucleic acid sequence (SEQ ID NO:34) for a triple cassette embodiment of the present invention.

Figure 24 shows the nucleic acid sequence (SEQ ID NO:35) for a hybrid gene expression cassette (TmManI-GnTI).

Figure 25 shows the nucleic acid sequence (SEQ ID NO:36) for the histone 3.1 promoter.

Figure 26 shows the nucleic acid sequence (SEQ ID NO:37) for the hybrid gene fusion 10 (TmManI-TmGnTI).

Figure 27 shows the nucleic acid sequence (SEQ ID NO:38) for the hybrid gene fusion TmManI-ManII (with the RbcS1 promoter sequence SEQ ID NO:39 shown).

Figure 28 shows the nucleic acid sequence (SEQ ID NO:39) for the RbcS1 promoter.

Figure 29 shows the nucleic acid sequence (SEQ ID NO:40) for the hybrid gene TmManI-15 ManII wherein the nucleic acid sequence (SEQ ID NO:33) encoding the transmembrane fragment is shown.

Figure 30 shows the nucleic acid sequence (SEQ ID NO:41) for the hybrid gene TmManI-GnTII.

Figure 31 shows the nucleic acid sequence (SEQ ID NO:42) for the Lhca promoter.

Figure 32 shows the nucleic acid sequence (SEQ ID NO:43) for the hybrid gene TmManI-GnTII wherein the nucleic acid sequence (SEQ ID NO:33) encoding the transmembrane fragment is shown

Figure 33 shows the nucleic acid sequence (SEQ ID NO:44) for the terminator sequence used (see below).

Figure 34 is a Western Blot which examines total protein glycosylation of plants of the present invention compared to control plants.

Figure 35 is a lectin blot with RCA on F1 progeny of crossed plants, said progeny made according to one embodiment of the present invention

Figure 36 is a Western Blot. Panel A was assayed with anti-IgG antibody. Panel B was assayed with an anti-HRP antibody. Panel C was assayed with a specific anti-Xyl antibody fraction. Panel D was assayed with a specific anti-Fucose antibody fraction. Panel E was assayed with the lectin RCA.

Figure 37 shows the nucleic acid sequence (SEQ ID NO:49) of a hybrid gene wherein the aminoterminal CTS region of an insect Mannosidase III gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

Figure 38 shows the corresponding amino acid sequence (SEQ ID NO:50) for the nucleic acid sequence of Figure 37.

Figure 39 shows the nucleic acid sequence (SEQ ID NO:51) of a hybrid gene wherein the aminoterminal CTS region of a human beta-1,4-galactosyltransferase (GalT) gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

5 Figure 40 shows the corresponding amino acid sequence (SEQ ID NO:52) for the nucleic acid sequence of Figure 39.

Figure 41 shows the nucleic acid sequence (SEQ ID NO:53) of a hybrid gene wherein the aminoterminal CTS region of an *Arabidopsis thaliana* GnTI gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

10 Figure 42 shows the corresponding amino acid sequence (SEQ ID NO:54) for the nucleic acid sequence of Figure 41.

Figure 43 shows the nucleic acid sequence (SEQ ID NO:55) of a hybrid gene wherein the aminoterminal CTS region of an *Arabidopsis thaliana* GnTII gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

15 Figure 44 shows the corresponding amino acid sequence (SEQ ID NO:56) for the nucleic acid sequence of Figure 43.

Figure 45 shows the nucleic acid sequence (SEQ ID NO:57) of a hybrid gene wherein the aminoterminal CTS region of a human beta-1,4-galactosyltransferase (GalT) gene is replaced by the CTS region of the human gene for GnTI.

20 Figure 46 shows the corresponding amino acid sequence (SEQ ID NO:58) for the nucleic acid sequence of Figure 45.

Figure 47 is a schematic of how enzymes might be localized to the Golgi.

Figure 48 is a non-limiting speculative schematic of how the "swapping" of regions of transferases might cause relocalization.

25

DETAILED DESCRIPTION OF THE INVENTION

Hybrid Enzymes

The nucleic acid sequences encoding the various glycosylation enzymes such as 30 mannosidases, GlcNAcTs, galactosyltransferases may be obtained using various recombinant DNA procedures known in the art, such as polymerase chain reaction (PCR) or screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) 35 and nucleic acid sequence-based amplification (NASBA) or long range PCR may be used.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if an amount of a portion of a gene or its specific RNA, or a fragment thereof, is available and can be purified and

labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe [Benton and Davis, *Science* 196:180 (1977); Grunstein and Hogness, *Proc. Natl. Acad. Sci. U.S.A.* 72:3961 (1975)]. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which-hybrid-select the proper mRNAs, can be selected which produce a protein that, e.g., has similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, or antigenic properties as known for the protein of interest.

A nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane portion of a first enzyme and a catalytic portion of a second enzyme may be obtained as follows. The sequence encoding the transmembrane portion is removed from the second enzyme, leaving a nucleic acid sequence comprising a nucleic acid sequence encoding the C-terminal portion of the second enzyme, which encompasses the catalytic site. The sequence encoding the transmembrane portion of the first enzyme is isolated or obtained via PCR and ligated to the sequence encoding a sequence comprising the C-terminal portion of the second enzyme.

15

Modified Enzymes

A nucleic acid sequence encoding a protein, particularly enzymes such as galactosyltransferases, mannosidases and N-acetylglucosamine transferases that are retained in the ER may be obtained by removing the sequence encoding the transmembrane fragment and substituting it for a methionine (initiation of translation) codon and by inserting between the last codon and the stop codon of galactosyltransferase the nucleic acid sequence encoding an ER retention signal such as the sequence encoding KDEL (amino acid residue sequence: lysine-aspartic acid-glutamic acid-leucine) [Rothman *Cell* 50:521 (1987)].

25 **Using Domains and Portions Thereof**

As noted above, the phrases "at least a portion of" or a "fragment of" refers to the minimal amino acid sequence necessary for a protein or a peptide to retain its natural or native function. For example, the function of an enzyme could refer to its enzymatic or catalytic role, its ability to anchor a protein in the Golgi apparatus, or as a signal peptide. Thus, the phrases "at least a portion of a transmembrane domain" or "a fragment of a transmembrane domain" each refer to the smallest amino acid segment of a larger transmembrane domain that still retains at least part of the native transmembrane functionality (for example, the function may be evident, albeit decreased). As another example, the phrases "at least a portion of a catalytic region" or "a fragment of a catalytic region" each refer to the smallest amino acid segment of a larger catalytic region that still retains at least part of the native catalytic functionality (again, even if somewhat decreased). As discussed herein, one skilled in the art will know the minimal amino acid segment that is necessary for a protein or a peptide to retain at least some of the functionality of the native protein or peptide.

The glycosyltransferase enzymes are typically grouped into families based on the type of sugar they transfer (galactosyltransferases, sialyltransferases, etc.). Based on amino-acid sequence similarity and the stereochemical course of the reaction, glycosyltransferases can be classified into at least 27 and perhaps as many as 47 different families [Campbell *et al.*, *Biochem. J.* 326:929-939 (1997), *Biochem. J.* 329:719 (1998)].

5 The majority of glycosyltransferases cloned to date are type II transmembrane proteins (*i.e.*, single transmembrane domain with the NH₂ terminus in the cytosol and the COOH terminus in the lumen of the Golgi apparatus). Regardless of how they are classified, all glycosyltransferases share some common structural features: a short NH₂-terminal cytoplasmic tail, a 16-20 amino acid signal-anchor or transmembrane domain, and an extended stem region
10 which is followed by the large COOH-terminal catalytic domain. The cytoplasmic tail appears to be involved in the specific localization of some types of glycosyltransferases to the Golgi [Milland *et al.*, *J. Biol. Chem.* 277:10374-10378]. The signal anchor domains can act as both uncleavable signal peptides and as membrane-spanning regions that orient the catalytic domains of the glycosyltransferases within the lumen of the Golgi apparatus.

15 In one embodiment of the present invention, a portion defined by the N-terminal 77 amino acids of *Nicotiana benthamiana* (tobacco) acetylglucosaminyltransferase I are contemplated for use in the hybrid enzyme(s), since this portion has been found to be sufficient to target to and to retain a reporter protein in the plant Golgi apparatus [Essl *et al.*, *FEBS Lett.* 453:169-173 (1999)].

Subcellular localization in tobacco of various fusion proteins between the putative cytoplasmic, 20 transmembrane and stem domains revealed that the cytoplasmic-transmembrane domains alone were sufficient to sustain Golgi retention of β 1,2-xylosyltransferase without the contribution of any luminal sequences [Dirnberger *et al.*, *Plant Mol. Biol.* 50:273-281 (2002)]. Thus, as noted above, certain embodiments of the present invention utilize portions of the CTS region which involve only the cytoplasmic-transmembrane domains (or portions thereof) without utilizing the stem region of 25 the CTS region. However, while some types of glycosyltransferases rely primarily on their transmembrane domain for Golgi retention, other types require their transmembrane region and sequences flanking one or both sides of this region [Colley, *Glycobiology* 7:1-13 (1997)]. For example, the N-terminal peptide encompassing amino acids 1 to 32 appears to be the minimal targeting signal sufficient to localize β 1,6 N-acetylglucosaminyltransferase to the Golgi. This 30 peptide makes up the cytoplasmic and transmembrane domains of this enzyme [Zerfaoui *et al.*, *Glycobiology* 12:15-24].

A great deal of information is available on the amino acid sequences of the domains for specific glycosyltransferases. For example, the amino acid sequence of the mammalian galactosyltransferase provided in GenBank Accession No. AAM17731 has the "stem" and 35 "catalytic" domains spanning residues 19 to 147 and residues 148 to 397, respectively [U.S. Patent No. 6,416,988, hereby incorporated by reference] – and the present invention, in certain embodiments, specifically contemplates such portions for use in the hybrid enzyme(s). The amino acid sequence of the rat liver sialyltransferase provided in GenBank Accession No. AAC91156 has a

9-amino acid NH₂-terminal cytoplasmic tail, a 17-amino acid signal-anchor domain, and a luminal domain that includes an exposed stem region followed by a 41 kDa catalytic domain [Hudgin *et al.*, *Can. J. Biochem.* 49:829-837 (1971); U.S. Patent Nos. 5,032,519 and 5,776,772, hereby incorporated by reference]. Known human and mouse β 1,3-galactosyltransferases have a catalytic domain with 5 eight conserved regions [Kolbinger *et al.*, *J. Biol. Chem.* 273:433-440 (1998); Hennet *et al.*, *J. Biol. Chem.* 273:58-65 (1998); U.S. Patent No. 5,955,282, hereby incorporated by reference]. For example, the amino acid sequence of mouse UDP-galactose: β -N-acetylglucosamine β 1,3-galactosyltransferase-I provided in GenBank Accession No. NM020026 has the following catalytic regions: region 1 from residues 78-83; region 2 from residues 93-102; region 3 from residues 116-119; region 4 from residues 147-158; region 5 from residues 172-183; region 6 from residues 203-206; region 7 from amino acid residues 236-246; and region 8 from residues 264-275. [Hennet *et al.*, *supra.*] – all of which are contemplated in certain embodiments of the present invention as useful portions in the context of the hybrid enzyme(s) discussed above.

While earlier comparisons amongst known cDNA clones of glycosyltransferases had 15 revealed very little sequence homology between the enzymes [Paulson *et al.*, *J. Biol. Chem.* 264:17615-618 (1989)], more recent advances have made it possible to deduce conserved domain structures in glycosyltransferases of diverse specificity [Kapitonov *et al.*, *Glycobiology* 9:961-978 (1999)]. For example, the nucleic acid and amino acid sequences of a number of 20 glycosyltransferases have been identified using sequence data provided by the complete genomic sequences obtained for such diverse organisms as *Homo sapiens* (humans), *Caenorhabditis elegans* (soil nematode), *Arabidopsis thaliana* (thale cress, a mustard) and *Oryza sativa* (rice).

As a result of extensive studies, common amino acid sequences have been deduced for 25 homologous binding sites of various families of glycosyltransferases. For example, sialyltransferases have sialyl motifs that appear to participate in the recognition of the donor substrate, CMP-sialic acid [Paulson *et al.*, *J. Biol. Chem.*, 264:17615-17618 (1989); Datta *et al.*, *J. Biol. Chem.*, 270:1497-1500 (1995); Katsutoshi, *Trends Glycosci. Glycotech.* 8:195-215 (1996)]. The hexapeptide RDKKND in Gal α 1-3 galactosyltransferase and RDKKNE in GlcNAc β 1-4 galactosyltransferase have been suggested as the binding site for UDP-Gal [(Joziasse *et al.*, *J. Biol. Chem.*, 260:4941-4951 (1985), *J. Biol. Chem.*, 264:14290-14297 (1989); Joziasse, *Glycobiology*, 2:271-277 (1992)].

30 A small, highly-conserved motif formed by two aspartic acid residues (DXD), which is frequently surrounded by a hydrophobic region, has been identified in a large number of different eukaryotic transferases, including α -1, 3-mannosyltransferase, β -1, 4-galactosyltransferases, α -1, 3-galactosyltransferases, glucuronyltransferases, fucosyltransferases, glycogenins and others [Wiggins *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 95:7945-7950 (1998)]. Mutation studies indicate that this motif 35 is necessary for enzymatic activity [Busch *et al.*, *J. Biol. Chem.* 273:19566-19572 (1998); Wang *et al.*, *J. Biol. Chem.* 277:18568-18573 (2002)]. Multiple peptide alignment showed several motifs corresponding to putative catalytic domains that are conserved throughout all members of the β 3-

galactosyltransferase family, namely, a type II transmembrane domain, a conserved DxD motif, an N-glycosylation site and five conserved cysteines [Gromova *et al.*, *Mol. Carcinog.* 32:61-72 (2001)].

Through the use of BLAST searches and multiple alignments, the E-X₇-E motif was found to be a highly conserved among the members of four families of retaining glycosyltransferases [Cid *et al.*, *J. Biol. Chem.* 275:33614-33621 (2000)]. The O-linked acetylglucosaminyltransferases (GlcNAc) add a single β-N-acetylglucosamine moiety to specific serine or threonine hydroxyls. BLAST analyses, consensus secondary structure predictions and fold recognition studies indicate that a conserved motif in the second Rossmann domain points to the UDP-GlcNAc donor-binding site [Wrabl *et al.*, *J. Mol. Biol.* 314:365-374 (2001)]. The β1, 3-glycosyltransferase enzymes identified to date share several conserved regions and conserved cysteine residues, all being located in the putative catalytic domain. Site-directed mutagenesis of the murine β3GatT-I gene (Accession No. AF029790) indicate that the conserved residues W101 and W162 are involved in the binding of the UDP-galactose donor, the residue W315 in the binding of the N-acetylglucosamine- β-p-nitrophenol acceptor, and the domain including E264 appears to participate in the binding of both substrates [Malissard *et al.*, *Eur. J. Biochem.* 269:233-239 (2002)].

Expression of Proteins of Interest in Plant Host System

The nucleic acid encoding the hybrid or modified enzymes or other heterologous proteins, such as a heterologous glycoprotein may be inserted according to certain embodiments of the present invention into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation, as well as selectable markers. These include but are not limited to a promoter region, a signal sequence, 5' untranslated sequences, initiation codon (depending upon whether or not the structural gene comes equipped with one), and transcription and translation termination sequences. Methods for obtaining such vectors are known in the art (see WO 01/29242 for review).

Promoter sequences suitable for expression in plants are described in the art, *e.g.*, WO 91/198696. These include non-constitutive promoters or constitutive promoters, such as, the nopaline synthetase and octopine synthetase promoters, cauliflower mosaic virus (CaMV) 19S and 35S promoters and the figwort mosaic virus (FMV) 35 promoter (see U.S. Pat. Nos. 5, 352,605 and 6,051,753, both of which are hereby incorporated by reference). Promoters used may also be tissue specific promoters targeted for example to the endosperm, aleurone layer, embryo, pericarp, stem, leaves, tubers, roots, and the like.

A signal sequence allows processing and translocation of a protein where appropriate. The signal can be derived from plants or could be non-plant signal sequences. The signal peptides direct the nascent polypeptide to the endoplasmic reticulum, where the polypeptide subsequently undergoes post-translational modification. Signal peptides can routinely be identified by those of skill in the art. They typically have a tripartite structure, with positively charged amino acids at the N-terminal end,

followed by a hydrophobic region and then the cleavage site within a region of reduced hydrophobicity.

The transcription termination is routinely at the opposite end from the transcription initiation regulatory region. It may be associated with the transcriptional initiation region or from a different gene and may be selected to enhance expression. An example is the NOS terminator from 5 *Agrobacterium* Ti plasmid and the rice alpha-amylase terminator. Polyadenylation tails may also be added. Examples include but are not limited to *Agrobacterium* octopine synthetase signal, [Gielen *et al.*, *EMBO J.* 3:835-846 (1984)] or nopaline synthase of the same species [Depicker *et al.*, *Mol. Appl. Genet.* 1:561-573 (1982)].

10 Enhancers may be included to increase and/or maximize transcription of the heterologous protein. These include, but are not limited to peptide export signal sequence, codon usage, introns, polyadenylation, and transcription termination sites (see WO 01/29242).

15 Markers include preferably prokaryote selectable markers. Such markers include resistance toward antibiotics such as ampicillin, tetracycline, kanamycin, and spectinomycin. Specific examples include but are not limited to streptomycin phosphotransferase (*spt*) gene coding for streptomycin resistance, neomycin phosphotransferase (*nptII*) gene encoding kanamycin or geneticin resistance, hygromycin phosphotransferase (*hpt*) gene encoding resistance to hygromycin.

20 The vectors constructed may be introduced into the plant host system using procedures known in the art (reviewed in WO 01/29242 and WO 01/31045). The vectors may be modified to intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *A. tumefaciens*. Alternatively, the vectors used in the methods of the present invention may be *Agrobacterium* vectors. Methods for introducing the vectors include but are not limited to microinjection, velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface and 25 electroporation. The vector may be introduced into a plant cell, tissue or organ. In a specific embodiment, once the presence of a heterologous gene is ascertained, a plant may be regenerated using procedures known in the art. The presence of desired proteins may be screened using methods known in the art, preferably using screening assays where the biologically active site is detected in such a way as to produce a detectable signal. This signal may be produced directly or indirectly.

30 Examples of such assays include ELISA or a radioimmunoassay.

Transient Expression

The present invention specifically contemplates both stable and transient expression of the above-described hybrid enzymes. Techniques for transforming a wide variety of higher plant species 35 for transient expression of an expression cassette are well known [see, for example, Weising *et al.*, *Ann. Rev. Genet.* 22:421-477(1988)]. Variables of different systems include type nucleic acid transferred (DNA, RNA, plasmid, viral), type of tissue transformed, means of introducing transgene(s), and conditions of transformation. For example, a nucleic acid construct may be

introduced directly into a plant cell using techniques ranging from electroporation, PEG poration, particle bombardment, silicon fiber delivery, microinjection of plant cell protoplasts or embryogenic callus or other plant tissue, or Agrobacterium-mediated transformation [Hiei *et al.*, *Plant J.* 6:271-282 (1994)]. Because transformation efficiencies are variable, internal standards (eg, 35S-Luc) are often used to standardize transformation efficiencies.

5 Expression constructs for transient assays include plasmids and viral vectors. A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

10 Plant tissues suitable for transient expression include cultured cells, either intact or as protoplasts (in which the cell wall is removed), cultured tissue, cultured plants, and plant tissue such as leaves.

15 Some transient expression methods utilize gene transfer into plant cell protoplasts mediated by electroporation or polyethylene glycol (PEG). These methods require the preparation and culture of plant protoplasts, and involve creating pores in the protoplast through which nucleic acid is transferred into the interior of the protoplast.

Exemplary electroporation techniques are described in Fromm *et al.*, *Proc. Natl. Acad. Sci.* 82: 5824 (1985). The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al.*, *EMBO J.* 3: 2717-2722 (1984). PEG-mediated transformation of tobacco protoplasts, which includes the steps of isolation, purification, and transformation of the 20 protoplasts, are described in Lyck *et al.*, (1997) *Planta* 202: 117-125 and Scharf *et al.*, (1998) *Mol Cell Biol* 18: 2240-2251, and Kirschner *et al.*, (2000) *The Plant J.* 24(3): 397-411. These methods have been used, for example, to identify cis-acting elements in promoters activated by external 25 stimuli, Abel and Theologis (1994) *Plant J.* 5: 421-427; Hattori *et al.*, (1992) *Genes Dev* 6: 609-618; Sablowski *et al.*, (1994) *EMBO J.* 13: 128-137; and Solano *et al.*, (1995) *EMBO J.* 14: 1773-1784), as well as for other gene expression studies (U. S. Patent 6,376,747, hereby incorporated by reference).

Ballistic transformation techniques are described in Klein *et al.*, (1987) *Nature* 327: 70-73. Biostatic transient transformation is used with suspension cells or plant organs. For example, it has been developed for use in *Nicotiana tabacum* leaves, Godon *et al.* (1993) *Biochimie* 75(7): 591-595. 30 It has also been used in investigating plant promoters, (Baum *et al.*, (1997) *Plant J.* 12: 463-469; Stromvik *et al.*, (1999) *Plant Mol Biol* 41(2): 217-31, Tuerck and Fromm (1994) *Plant Cell* 6: 1655-1663; and U. S. Patent 5,847,102, hereby incorporated by reference), and to characterize transcription factors (Goff *et al.*, (1990) *EMBO J.* 9: 2517-2522; Gubler *et al.*, (1999) *Plant J.* 17: 1-9; and Sainz *et al.*, (1997) *Plant Cell* 9: 611-625).

35 Other methods allow visualization of transient expression of genes *in situ*, such as with onion epidermal peels, in which GFP expression in various cellular compartments was observed (Scott *et al.*, (1999) *Biotechniques* 26(6): 1128-1132

Nucleic acids can also be introduced into plants by direct injection. Transient gene expression can be obtained by injection of the DNA into reproductive organs of a plant (see, for example, Pena *et al.*, (1987) *Nature*, 325:274), such as by direct DNA transfer into pollen (see, for example, Zhou *et al.*, (1983) *Methods in Enzymology*, 101:433; D. Hess (1987) *Intern Rev. Cytol.*, 107:367; Luo *et al.*, (1988) *Plant Mol. Biol. Reporter*, 6:165. DNA can also be injected directly into the cells of immature embryos (see, for example, Neuhaus *et al.*, (1987) *Theor. Appl. Genet.* 75:30; and Benbrook *et al.*, (1986) in *Proceedings Bio Expo* 1986, Butterworth, Stoneham, Mass., pp. 27-54).

Agrobacterium-mediated transformation is applicable to both dicots and monocots.

Optimized methods and vectors for Agrobacterium-mediated transformation of plants in the family Graminae, such as rice and maize have been described (see, for example, Heath *et al.*, (1997) *Mol. Plant-Microbe Interact.* 10:221-227; Hiei *et al.*, (1994) *Plant J.* 6:271-282 and Ishida *et al.*, (1996) *Nat. Biotech.* 14:745-750). The efficiency of maize transformation is affected by a variety of factors including the types and stages of tissue infected, the concentration of Agrobacterium, the tissue culture media, the Ti vectors and the maize genotype.

Another useful basic transformation protocol involves a combination of wounding by particle bombardment, followed by use of Agrobacterium for DNA delivery (see, for example, Bidney *et al.*, (1992) *Plant Mol. Biol.* 18:301-313). Both intact meristem transformation and a split meristem transformation methods are also known (U. S. Patent 6,300,545, hereby incorporated by reference).

Additional methods utilizing Agrobacteria include agroinfection and agroinfiltration. By inserting a viral genome into the T-DNA, Agrobacterium can be used to mediate the viral infection of plants (see, for example, U. S. Patent 6,300,545, hereby incorporated by reference). Following transfer of the T-DNA to the plant cell, excision of the viral genome from the T-DNA (mobilization) is required for successful viral infection. This Agrobacterium-mediated method for introducing a virus into a plant host is known as agroinfection (see, for example, Grimsley, "Agroinfection" pp. 325-342, in *Methods in Molecular Biology*, vol 44: Agrobacterium Protocols, ed. Gartland and Davey, Humana Press, Inc., Totowa, N.J.; and Grimsley (1990) *Physiol. Plant.* 79:147-153).

The development of plant virus gene vectors for expression of foreign genes in plants provides a means to provide high levels of gene expression within a short time.

Suitable viral replicons include double-stranded DNA from a virus having a double stranded DNA genome or replication intermediate. The excised viral DNA is capable of acting as a replicon or replication intermediate, either independently, or with factors supplied in trans. The viral DNA may or may not encode infectious viral particles and furthermore may contain insertions, deletions, substitutions, rearrangements or other modifications. The viral DNA may contain heterologous DNA, which is any non-viral DNA or DNA from a different virus. For example, the heterologous DNA may comprise an expression cassette for a protein or RNA of interest.

Super binary vectors carrying the vir genes of Agrobacterium strains A281 and A348 are useful for high efficiency transformation of monocots. However, even without the use of high

efficiency vectors, it has been demonstrated that T-DNA is transferred to maize at an efficiency that results in systemic infection by viruses introduced by agroinfection, although tumors are not formed (Grimsley *et al.*, (1989) *Mol. Gen. Genet.* 217:309-316). This is because integration of the T-DNA containing the viral genome is not required for viral multiplication, since the excised viral genome

5 acts as an independent replicon.

Another Agrobacteria-mediated transient expression assay is based on Agrobacterium-mediated transformation of tobacco leaves in planta (Yang *et al.*, (2000) *The Plant J.* 22(6): 543-551). The method utilizes infiltration of agrobacteria carrying plasmid constructs into tobacco leaves, and is referred to as agroinfiltration; it has been utilized used to analyze *in vivo* expression of 10 promoters and transcription factors in as little as 2-3 days. It also allows examination of effects of external stimuli such as pathogen infections and environmental stresses on promoter activity *in situ*.

Example 1

An *Arabidopsis thaliana* cDNA encoding β 1,2-xylosyltransferase was isolated from a cDNA library by a previously described PCR based sibling selection procedure [Bakker *et al.*, *BBRC* 261:829 (1999)]. Xylosyltransferase activity was confirmed by immunostaining of transfected CHO cells with a xylose specific antibody purified from rabbit-anti-horseradish-peroxidase antiserum. A DNA fragment covering the N-terminal part of the xylosyltransferase was amplified using primers:

20 XylTpvuF:ATACTCGAGTTAACAAATGAGTAAACGGAATC (SEQ ID NO:45)
and XylTpvuR:TTCTCGATGCCGATTGGTTATT (SEQ ID NO:46)
XhoI and HpaI restriction sites were introduced in front of the start codon and a PvuI was introduced at the reverse end. A C-terminal fragment from Human β 1,4galactosyltransferase (acc.no. x55415, Aoki 1992) was amplified using primers GalTpvuF:GCCGCCGCGATCGGGCAGTCCTCC (SEQ 25 ID NO:47) and GalTrev: AACGGATCCACGCTAGCTCGGTGTCCCGAT (SEQ ID NO:48) thus introducing PvuI and BamHI sites. The XhoI/PvuI and PvuI/BamHI digested PCR fragments were ligated in XhoI/BamHI digested pBluescriptSK+ and sequenced. The resulting open reading frame encodes a fusion protein containing the first 54 amino acids of *A. thaliana* β 1,2-xylosyltransferase fused with amino acid 69 to 398 of human β 1,4galactosyltransferase and is designated as TmXyl-GalT. The fragment was cloned into a plant expression vector between the CaMV35S promoter and 30 Nos terminator, using HpaI/BamHI. The clone was introduced into *Nicotiana tabacum* (samsun NN) as described for native human β 1,4galactosyltransferase [Bakker *et al.*, *Proc. Nat. Acad. Sci. USA* 98:2899 (2001)].

Protein extract of transgenic plants and Western Blots were made as described [Bakker *et al.*, 35 *Proc. Nat. Acad. Sci. USA* 98:2899 (2001)]. Based on reaction with the lectin RCA, a transgenic plant expressing TmXylGalT was selected for further glycan analysis by MALDI-TOF [Elbers *et al.*, *Plant Physiology* 126:1314 (2001) and compared with glycans isolated from plants expressing native β 1,4galactosyltransferase and with glycans from wild-type plants. Relative peak areas of the

MALDI-TOF spectrum are given in Table 1. That is to say, Table 1 is a comparison of the results of mass spec (MALDI-TOF) analysis of N-glycans of endogenous glycoproteins of control tobacco ("Tobacco"), transgenic tobacco expressing human beta-1,4-galactosyltransferase ("GalT") and transgenic tobacco plants expressing the beta-1,4-galactosyltransferase gene of which the CTS region has been replaced with that of beta-1,2-xylosyltransferase ("TmXyl-GalT").

TABLE 1					
m/z	Type	Tobacco	GalT	TmXyl-GalT	
933	M3		3	7	
1065	XM3	10	16	3	
1079	FM3			4	
1095	M4			9	
1211	FXM3	41	27		
1257	M5	4	5	23	
1268	GNXM3		4		
1298	GalGNM3			6	
1298	GNM4				
1414	GNFXM3	27	13	5	
1419	M6	7	8	10	
1460	GalGNM4			11	
1460	GNM5				
1485	GN2FM3		4		
1576	GalGNFXM3		5		
1576	GNFXM4				
1581	M7	3		4	
1606	GNFM5			3	
1606	GalGNFM4				
1617	GN2FXM3	8	9		
1622	GalGNM5			9	
1622	GNM6				
1743	M8		2	3	
1768	GalGNFM5			3	
1768	GNFM6				
1779	GalIGN2FXM3		2		
1905	M9			1	
1941	Gal2GN2FXM3		2		
	TOTAL	100	100	101	

These data show that:

1. In TmXylGalT plants, xylosylation and fucosylation of the glycans is dramatically reduced: 82% of the glycans do not carry xylose nor fucose as compared to 14% in wild-type plants.
2. Galactosylation has increased from 9% in GalT plants to 32% in TmXylGalT plants.

Example 2

A transgenic plant expressing said TmXyl-GalT gene (TmXyl-GalT-12 plant) was selected (above) based on lectin blotting using biotin-labelled RCA (Vector Laboratories, Burlingame, California). Comparison of protein extracts of MGR48 transgenic (control) plant, a selected 5 transgenic plant expressing the unmodified human β 1,4-galactosyltransferase gene and TmXyl-GalT-12 plant for the presence of xylose and fucose using anti-HRP (horseradish peroxidase) polyclonal antibody (known for high anti-xylose and anti-fucose reactivity) clearly showed reduced xylose and fucose (Figure 34: "Anti-HRP"). Western blotting using an anti-xylose fraction of the 10 anti-HRP and an anti-fucose fraction (each of which can be prepared by affinity chromatography over the appropriate ligand) showed that especially xylose was reduced compared to control plants (Figure 34: anti-Fuc" and "anti-Xyl").

Example 3

The TmXyl-GalT-12 plant was crossed with a transgenic plant expressing the monoclonal 15 antibody MGR48 from a single T-DNA integration event (MGR48-31) and which was first made homozygous by selecting offspring plants not segregating for the kanamycin resistance marker and antibody production (MGR48-31-4). Pollen of MGR48-31-4 was used for pollination of emasculated TmXyl-GalT-12 plants. Vice versa, pollen of TmXyl-GalT-12 plant was used for fertilization on 20 emasculated MGR48-31-4 plants. A number of F1 plants were analyzed for the presence of MGR48 by western blotting and for galactosylation of endogenous glycoproteins by lectin blotting using RCA (Figure 35). One plant expressing MGR48 and showing galactosylation of endogenous 25 glycoproteins was selected for further analysis. This plant was identified as XGM8.

Seeds from TmXyl-GalT-12 (♀) x MGR48-31-4 (♂) were sown and F1 offspring plants (XGM) were analysed for antibody production by Western blotting and for galactosylation by lectin 30 blotting using biotinylated RCA120 (Vector Labs., Burlingame, California) using standard techniques as described before. All plants as expected expressed the monoclonal antibody MGR48 and the majority also had galactosylated glycans as depicted from lectin blotting using RCA120. A single plant expressing both antibody MGR48 and having galactosylated N-glycans was chosen for further analysis (XGM8) (TmXyl-GalT-12 X MGR48-31-4 offspring plant 8). The monoclonal recombinant MGR48 antibody was purified from this plant as described before and submitted to N-glycan analysis by MALDI-TOF.

Briefly, XGM8 plant was grown in greenhouse for antibody production under optimal conditions [Elbers *et al.*, *Plant Physiology* 126:1314 (2001)]. Protein extract of leaves of transgenic XGM8 plant was made and monoclonal antibody was purified using protein G chromatography as 35 described [Bakker *et al.*, *Proc. Nat. Acad. Sci. USA* 98:2899 (2001)]. MALDI-TOF of N-glycans of purified monoclonal antibody was as described (Elbers *et al.*, 2001, *supra*). The presence of galactose on glycans was established by enzyme sequencing using bovine testis β -galactosidase as described (Bakker *et al.*, 2001, *supra*; Table 2). Table 2 (below) is a comparison of the results of

mass spec (MALDI-TOF) analysis of N-glycans of endogenous glycoproteins ("Xyl-GalT Endo") of a F1 hybrid of TmXyl-GalT-12 plant and plant producing rec-mAb (MGR48) and of N-glycans of rec-mAb purified by protein G chromatography from said F1 hybrid.

TABLE 2			Xyl-GalT Endo	Xyl-GalT IgG
m/z	Type			
933	M3		6	4
1065	XM3		2	2
1079	FM3		2	3
1095	M4		5	5
1136	GNM3		1	2
1211	FXM3		6	3
1241	FM4		3	2
1257	M5		17	12
1268	GNXM3		1	2
1282	GNFM3		2	3
1298	GalGNM3		3	4
1403	FM5		4	3
1414	GNFXM3		2	4
1419	M6		5	4
1430	GNXM4		2	2
1430	GalGNXM3			
1444	GNFM4		1	3
1444	GalGNFM3			
1460	GalGNM4		8	10
1460	GNM5			
1471	GN2XM3		1	
1485	GN2FM3		1	1
1501	GalGN2M3		1	1
1576	GalGNFXM3		2	3
1576	GNFXM4			
1581	M7		2	2
1593	GalGNXM4		1	2
1593	GNXM5			
1606	GNFM5		3	4
1606	GalGNFM4			
1617	GN2FXM3		2	1
1622	GalGNM5		6	6
1622	GNM6			
1647	GalGN2FM3		1	1
1663	Gal2GN2M3		1	1
1738	GNFXM5		1	2
1738	GalGNFXM4			
1743	M8		1	2
1754	GalGNXM5		1	2
1768	GalGNFM5		2	3
1768	GNFM6			

1784	GNM7	1	1
1784	GalGNM6		
1809	Gal2GN2FM3	2	1
1900	GNFXM6	1	
1900	GalGNFXM5		
1905	M9	1	1
	TOTAL	101	102

These data show that:

1. In the F1 hybrid, xylosylation and fucosylation of the glycans is dramatically reduced: 43% of the glycans of endogenous glycoproteins lack xylose and fucose as compared to only 14% in wild-type tobacco plants.
- 5 2. The glycans of purified mAb of this F1 hybrid have reduced xylose and fucose, 47% compared to 14% for wildtype tobacco. See also Figure 36, panels B-D.
3. Galactosylation of endogenous glycoproteins of F1 hybrid has increased from 9% in GalT plants to 37% in F1 TmXyl-GalT X MGR48 plant. See also Figure 35.
- 10 4. Purified rec-mAb from said F1 (see Figure 36, panel A) shows increased galactosylation; that is to say, 46% has galactose. See also Figure 36, panel E.

It should however be noted that the observed quantities (MALDI-TOF) do not necessarily reflect the molar ratios of said glycoforms in vivo. Quantification based on MALDI-TOF can be under- or overestimated depending on the specific glycoform under study. Also, since there is no molecular weight difference between Gal and Man, some peaks can not be annotated unambiguously unless there are clear differences in relative height of specific molecules before and after galactosidase treatment.

Example 4

20 A more direct comparison of xylose, fucose and galactose content was done by examining the MGR48 IgG antibodies from hybridoma, transgenic tobacco and TmXyl-GalT transgenic tobacco. As mentioned above, the TmXyl-GalT-12 plant was crossed with tobacco plant expressing MGR48 IgG (MGR48 tobacco) resulting in an F1 hybrid harbouring MGR48 TmXyl-GalT. An F1 plant was chosen for extraction and purification of MGR48 IgG. Antibodies from said plants (tobacco and TmXyl-GalT) were isolated and purified using protein G chromatography (Elbers *et al.*, 2001. *Plant Physiology* 126: 1314-1322). 300 nanograms amounts of each, hybridoma MGR48 and plant-derived recMGR48, were loaded on precast 12% SDS-PAGE gels (BioRad) and run. The contents of each lane were as follows: Lane 1, MGR48 from hybridoma; Lane 2, purified recMGR48 from normal transgenic tobacco plant; and Lane 3, purified recMGR48 from TmXyl-GalT transgenic plant. Following SDS-PAGE proteins were transferred to nitrocellulose using CAPS buffer. Blots were incubated with A, anti-mouse IgG; B, polyclonal rabbit anti-HRP (anti-xylose/(alpha 1,3-fucose); C, anti-xylose; D, anti-(alpha 1,3-) fucose antibodies; and E, biotinylated RCA. Detection

was with LumiLight on Lumi Imager following incubation with HRP-labelled sheep anti-mouse (panel A) or goat-anti-rabbit (panels B-D) antibodies and HRP-labeled streptavidin (E).

Panel A shows that approximately similar amounts of the MGR48 IgG was loaded for all lanes (1-3). L refers to Light chain and H, heavy chain of MGR48 IgG.

5 Panel B shows that the heavy chain of MGR48 antibody in lane 2 (tobacco) strongly reacts with anti-HRP as expected, whereas the heavy chain of hybridoma derived MGR48 (lane 1) does not (as expected). Hybridoma derived antibodies do not carry xylose and alpha 1, 3-fuctose residues.

10 Remarkably, MGR48 antibodies from TmXyl-GalT tobacco plant also do not react, suggesting that the heavy chain of antibody from this plant have significantly reduced (perhaps by 90% or more) the amounts of xylose and fucose residues on the N-glycans. This is confirmed by experiments depicted in panels C (anti-xylose) and D (anti-fucose). Panel E shows that the heavy chain of MGR48 antibody of hybridoma (lane 1) has a galactosylated N-glycan, whereas tobacco-derived MGR48 (lane 2) has not, both as expected. Heavy chain of MGR48 from the TmXyl-GalT plant (lane 3) also has galactosylated N-glycan due to the presence of the construct expressing the hybrid enzyme.

15 These data are in agreement with the data obtained from similar experiments using total protein extracts from similar plants (tobacco and TmXyl-GalT-12 plant) as shown previously and confirm that the novel trait introduced in tobacco from expression of TmXyl-GalT gene can be stably transmitted to offspring and a recombinant monoclonal antibody.

20 Example 5

Further characterization of the above-described F1 hybrid was performed by treatment with beta-galactosidase. Table 3 is a comparison of the results of mass spec (MALDI-TOF) analysis of N-glycans of rec-mAbs purified by protein G chromatography from an F1 hybrid of TmXyl-GalT and MGR48 plant before and after treatment of the glycans with beta-galactosidase.

25

TABLE 3		Type	Xyl-GalT IgG-	Xyl-GalT IgG+beta-galactosidase
m/z				
933		M3	4	4
1065		XM3	2	2
1079		FM3	3	3
1095		M4	5	4
1136		GNM3	2	3
1211		FXM3	3	4
1241		FM4	2	2
1257		M5	12	13
1268		GNXM3	2	3
1282		GNFM3	3	3
1298		GalGNM3	4	4
1403		FM5	3	2
1414		GNFXM3	4	5
1419		M6	4	3

1430	GNXM4	2	2
1430	GalGNXM3		
1444	GNFM4	3	3
1444	GalGNFM3		
1460	GalGNM4	10	14
1460	GNM5		
1471	GN2XM3		1
1485	GN2FM3	1	1
1501	GalGN2M3	1	
1576	GalGNFXM3	3	3
1576	GNFXM4		
1581	M7	2	2
1593	GalGNXM4	2	2
1593	GNXM5		
1606	GNFM5	4	6
1606	GalGNFM4		
1617	GN2FXM3	1	1
1622	GalGNM5	6	1
1622	GNM6		
1647	GalGN2FM3	1	
1663	Gal2GN2M3	1	
1738	GNFXM5	2	2
1738	GalGNFXM4		
1743	M8	2	2
1754	GalGNXM5	2	1
1768	GalGNFM5	3	1
1768	GNFM6		
1784	GNM7	1	1
1784	GalGNM6		
1809	Gal2GN2FM3	1	
1900	GNFXM6		1
1900	GalGNFXM5		
1905	M9	1	1
	TOTAL	102	100

These data show that:

1. Rec-mAbs from F1 hybrid contain galactose which can be deduced from the observed reduction of specific (galactose-containing) glycoforms after beta-galactosidase treatment and increase of glycoforms lacking galactose. Note the observed reduction of m/z 1622 from 6 to 1% and simultaneous increase of m/z 1460 from 10 to 14% which is the result of the removal of galactose from GalGNM5 to give rise to GNM5. The same is true for m/z 1768 (3 to 1% decrease) and corresponding m/z 1606 peak (4 to 6% increase). See also Figure 36, panel E.

- 5
2. Similarly a number of peaks that can be attributed to galactose containing glycans vanish upon treatment with galactosidase, especially m/z 1501, 1647 and 1663 confirming the presence of galactose.

Example 6

In another embodiment, the aminoterminal CTS region of an insect Mannosidase III gene (accession number: AF005034; mistakenly annotated as a Mannosidase II gene!) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 37).
5 The signal peptide sequence encodes a fully active signal peptide normally present at the aminotermminus of IgG sequences and has been used successfully in plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment
10 for ER retention. The hybrid Mannosidase III protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 7

In another embodiment, the aminoterminal CTS region of the human beta-1,4-galactosyltransferase (GalT) gene (accession A52551) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 39). The signal peptide sequence encodes a fully active signal peptide normally present at the aminotermminus of IgG sequences and has been used successfully in plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the
20 carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid beta-1,4-galactosyl-transferase protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 8

In another embodiment, the aminoterminal CTS region of *Arabidopsis thaliana* GnTI (acc. AJ243198) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 41). The signal peptide sequence encodes a fully active signal peptide normally present at the aminotermminus of IgG sequences and has been used successfully in plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic
30 reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid GnTI protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 9

In another embodiment, the aminoterminal CTS region of an *Arabidopsis thaliana* GnTII (acc. AJ249274) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 43). The signal peptide sequence encodes a fully active signal peptide normally present at the aminotermminus of IgG sequences and has been used successfully in

plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid GnTII protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

5

Example 10

In another embodiment, the aminoterminal CTS region of the human gene for beta-1,4-galactosyltransferase (GalT) gene is replaced by the CTS region of the human gene for GnTI (TmhuGnTI-GalT) (see Figure 45).

10

It is understood that the present invention is not limited to any particular mechanism. Nor is it necessary to understand the mechanism in order to successfully use the various embodiments of the invention. Nonetheless, it is believed that there is a sequential distribution of Golgi enzymes (Figure 47) and that the swapping in of transmembrane domains of plant glycosyltransferases causes 15 relocalization (Figure 48).

It is understood that the present invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described herein, as these may vary. It is also to be understood that the terminology used herein is used for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention. It must be noted that 20 as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

The invention described and claimed herein is not to be limited in scope by the specific 25 embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

30

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

WHAT IS CLAIMED IS:

1. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase.

5

2. The nucleic acid of Claim 1, wherein said plant glycosyltransferase is a xylosyltransferase.

3. The nucleic acid of Claim 1, wherein said plant glycosyltransferase is a N-acetylglucosaminyltransferase.

10

4. The nucleic acid of Claim 1, wherein said plant glycosyltransferase is a fucosyltransferase.

5. The nucleic acid of Claim 1, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

15

6. The nucleic acid of Claim 5, wherein said human galactosyltransferase is encoded by at least a portion of the nucleic acid sequence of SEQ ID NO:1.

20

7. An expression vector, comprising the nucleic acid of Claim 1.

8. A host cell transfected with the vector of Claim 7.

9. The host cell of Claim 8, wherein said host cell is a plant cell.

25

10. A cell suspension comprising the host cell of Claim 9.

11. The hybrid enzyme expressed by the plant cell of Claim 9.

12. The plant comprising the host cell of Claim 9.

30

13. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase.

35

14. The nucleic acid of Claim 13, wherein said first glycosyltransferase comprises a plant glycosyltransferase

15. The nucleic acid of Claim 14, wherein said plant glycosyltransferase is a xylosyltransferase.

16. The nucleic acid of Claim 14, wherein said plant glycosyltransferase is a fucosyltransferase.

17. The nucleic acid of Claim 13, wherein said second glycosyltransferase comprises a mammalian glycosyltransferase.

5 18. The nucleic acid of Claim 17, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

10 19. The nucleic acid of Claim 13, wherein said first glycosyltransferase comprises a first mammalian glycosyltransferase and said second glycosyltransferase comprises a second mammalian glycosyltransferase.

20. The nucleic acid of Claim 19, wherein said first mammalian glycosyltransferase is a non-human glycosyltransferase.

15 21. The nucleic acid of Claim 19, wherein said second mammalian glycosyltransferase is a human glycosyltransferase.

22. A method, comprising:

20 a. providing: i) a plant cell, and ii) an expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase; and b. introducing said expression vector into said plant cell under conditions such that said hybrid enzyme is expressed.

25 23. The method of Claim 22, wherein said plant glycosyltransferase is a xylosyltransferase.

24. The method of Claim 23, wherein said plant glycosyltransferase is a N-acetylglucosaminyltransferase.

30 25. The method of Claim 23, wherein said plant glycosyltransferase is a fucosyltransferase.

26. The method of Claim 22, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

35 27. The nucleic acid of Claim 26, wherein said human galactosyltransferase is encoded by at least a portion of the nucleic acid sequence of SEQ ID NO:1.

28. A method, comprising:

- a. providing: i) a plant cell, ii) a first expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; and
- b. introducing said first and second expression vectors into said plant cell under conditions such that said hybrid enzyme and said heterologous protein are expressed.

5

29. The method of Claim 28, wherein said heterologous protein is an antibody or antibody
10 fragment.

15

30. A method, comprising:

- a) providing: i) a first plant comprising a first expression vector, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, and ii) a second plant comprising a second expression vector, said second vector comprising nucleic acid encoding a heterologous protein; and
- b) crossing said first plant and said second plant to produce progeny expressing said hybrid enzyme and said heterologous protein.

20

31. A plant, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, said second vector comprising nucleic acid encoding a heterologous protein.

25

32. The plant of Claim 31, wherein said heterologous protein displays reduced amounts of fucose as compared to when the heterologous protein is expressed in a plant in the absence of said
30 hybrid enzyme

35

33. The plant of Claim 31, wherein the heterologous protein displays reduced amounts of xylose as compared to when the heterologous protein is expressed in a plant in the absence of said hybrid enzyme.

34. The plant of Claim 31, wherein the heterologous protein displays both reduced fucose and xylose, as compared to when the heterologous protein is expressed in a plant in the absence of said hybrid enzyme.

35. The plant of Claim 31, wherein the heterologous protein displays complex type bi-antennary glycans and contains galactose residues on at least one of the arms.

36. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a modified
5 mammalian glycosyltransferase, wherein a transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted.

37. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region or
portion thereof of a plant glycosyltransferase and a catalytic region of a mammalian
10 glycosyltransferase, wherein said CTS region is from a N-acetylglucosaminyltransferase I (GnTI) and said catalytic region is from a mannosidase II (ManII).

38. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region or
portion thereof of a plant glycosyltransferase and a catalytic region of a mammalian
15 glycosyltransferase, wherein said CTS region or portion thereof is from a
N-acetylglucosaminyltransferase I (GnTI) and said catalytic region is from a
N-acetylglucosaminyltransferase II (GnTII).

39. A plant host system, comprising (a) a nucleic acid sequence encoding a Mannosidase III
20 glycosyltransferase; (b) a nucleic acid sequence encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region or portion thereof of a plant glycosyltransferase and a catalytic domain of a mammalian glycosyltransferase

40. A method, comprising (a) introducing into said plant host system a vector comprising (i) a
25 nucleic acid sequence encoding a hybrid enzyme comprising a catalytic region of a galactosyltransferase not normally found in a plant and a transmembrane region of a protein, (ii) a nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane region of a N-acetylglucosaminyltransferase I (GnTI) and a catalytic region of a mannosidase II (ManII), (iii) a nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane region of an N-
30 acetylglucosaminyltransferase I (GnTI) and a catalytic region of a N-acetylglucosaminyltransferase II (GnTII); and (b) isolating a plant or portion thereof expressing said nucleic acid sequences.

41. A method, comprising:

- a. providing: i) a host cell, and ii) an expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase; and
- b. introducing said expression vector into said host cell under conditions such that said hybrid enzyme is expressed.

42. The method of Claim 41, wherein said first glycosyltransferase comprises a plant glycosyltransferase.
43. The method of Claim 42, wherein said plant glycosyltransferase is a xylosyltransferase.
44. The method of Claim 42, wherein said plant glycosyltransferase is a N-
5 acetylglucosaminyltransferase.
45. The method of Claim 42, wherein said plant glycosyltransferase is a fucosyltransferase.
46. The method of Claim 41, wherein said second glycosyltransferase comprises a mammalian glycosyltransferase.
47. The method of Claim 46, wherein said mammalian glycosyltransferase is a human
10 galactosyltransferase.
48. A method, comprising:
- a. providing: i) a host cell, ii) a first expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase, and iii) a
15 second expression vector comprising nucleic acid encoding a heterologous glycoprotein; and
- b. introducing said first and second expression vectors into said host cell under conditions such that said hybrid enzyme and said heterologous protein are expressed.
- 20 49. The method of Claim 48, wherein said heterologous protein is an antibody or antibody fragment.
50. The method of Claim 48, further comprising the step of c) isolating said heterologous protein.
51. The isolated heterologous protein produced according to the method of Claim 50.
- 25 52. A host cell, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a first glycosyltransferase and at least a portion of a catalytic region of a second glycosyltransferase, said second vector comprising nucleic acid encoding a heterologous protein.
- 30 53. The heterologous protein isolated from the host cell of Claim 52.

Common Plants Mammals

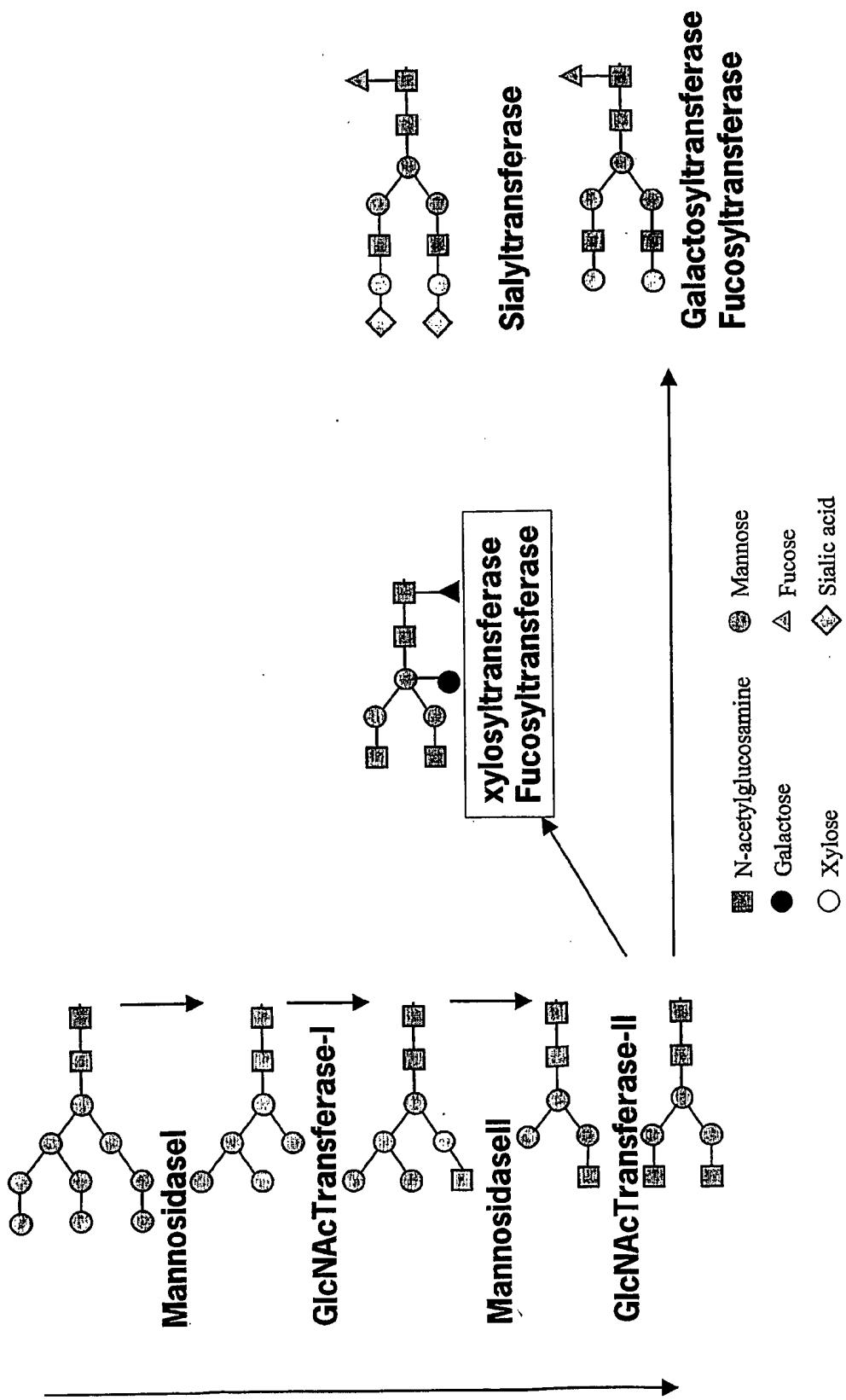


FIG. 1

2/57

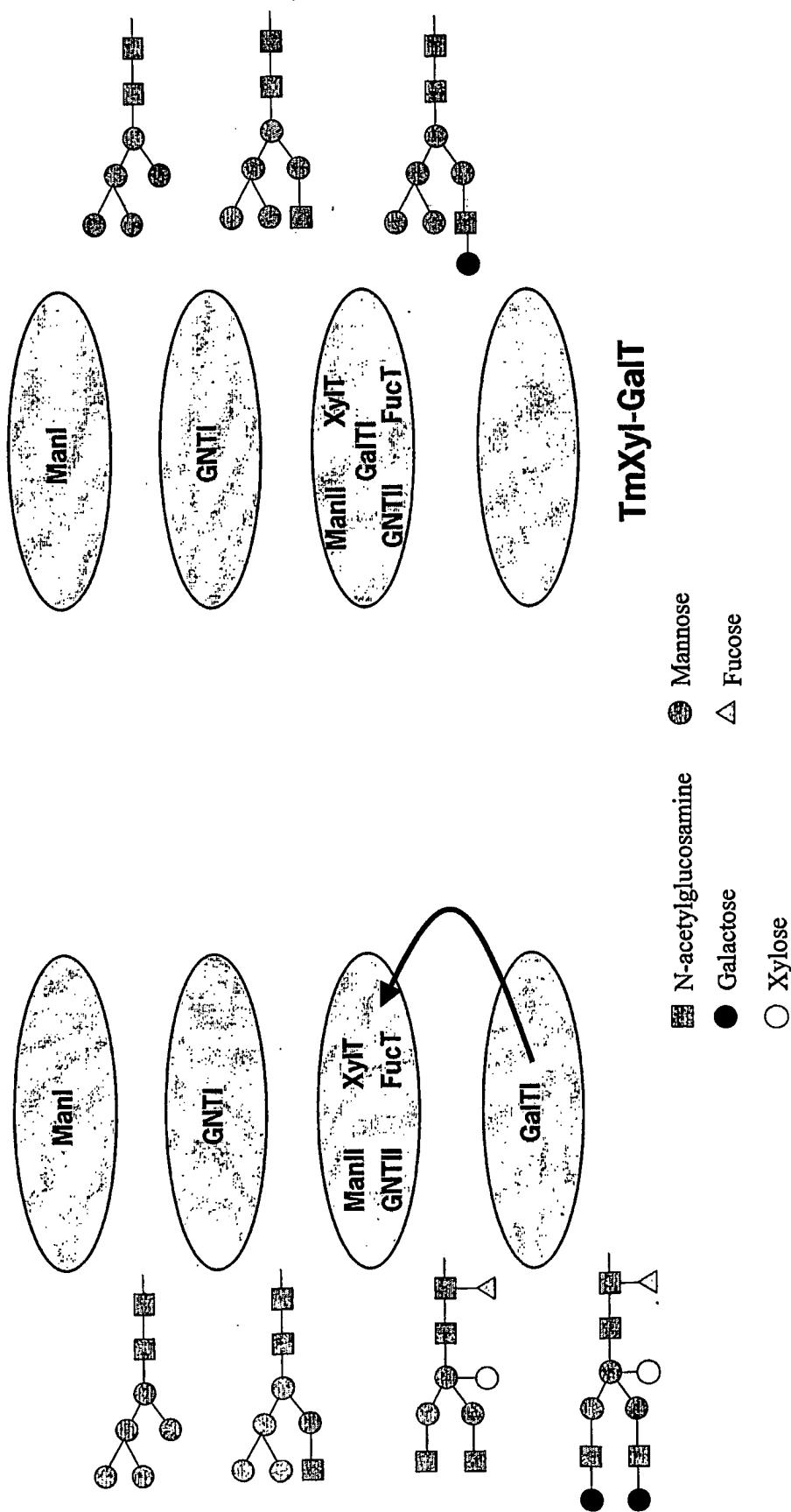
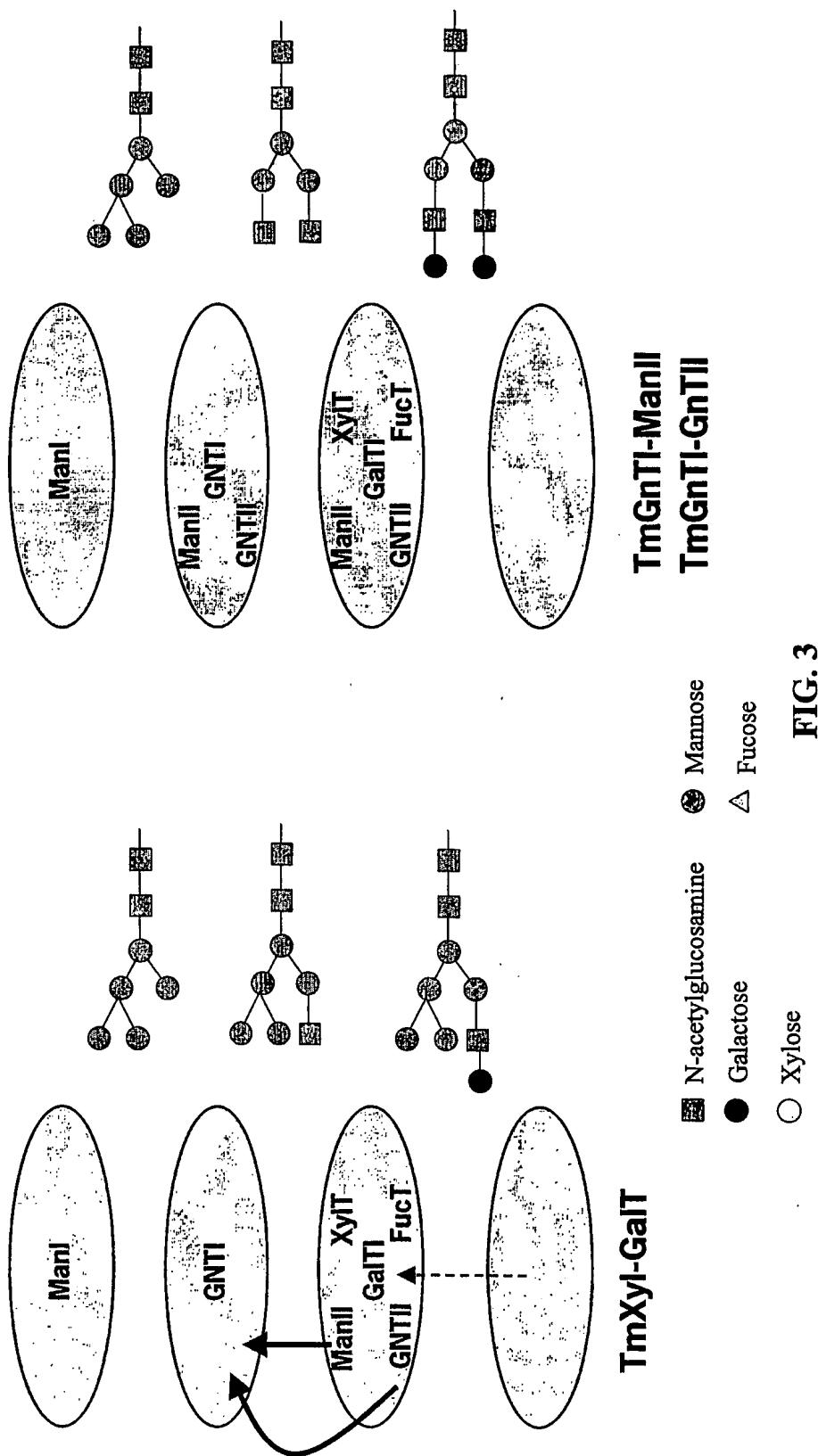
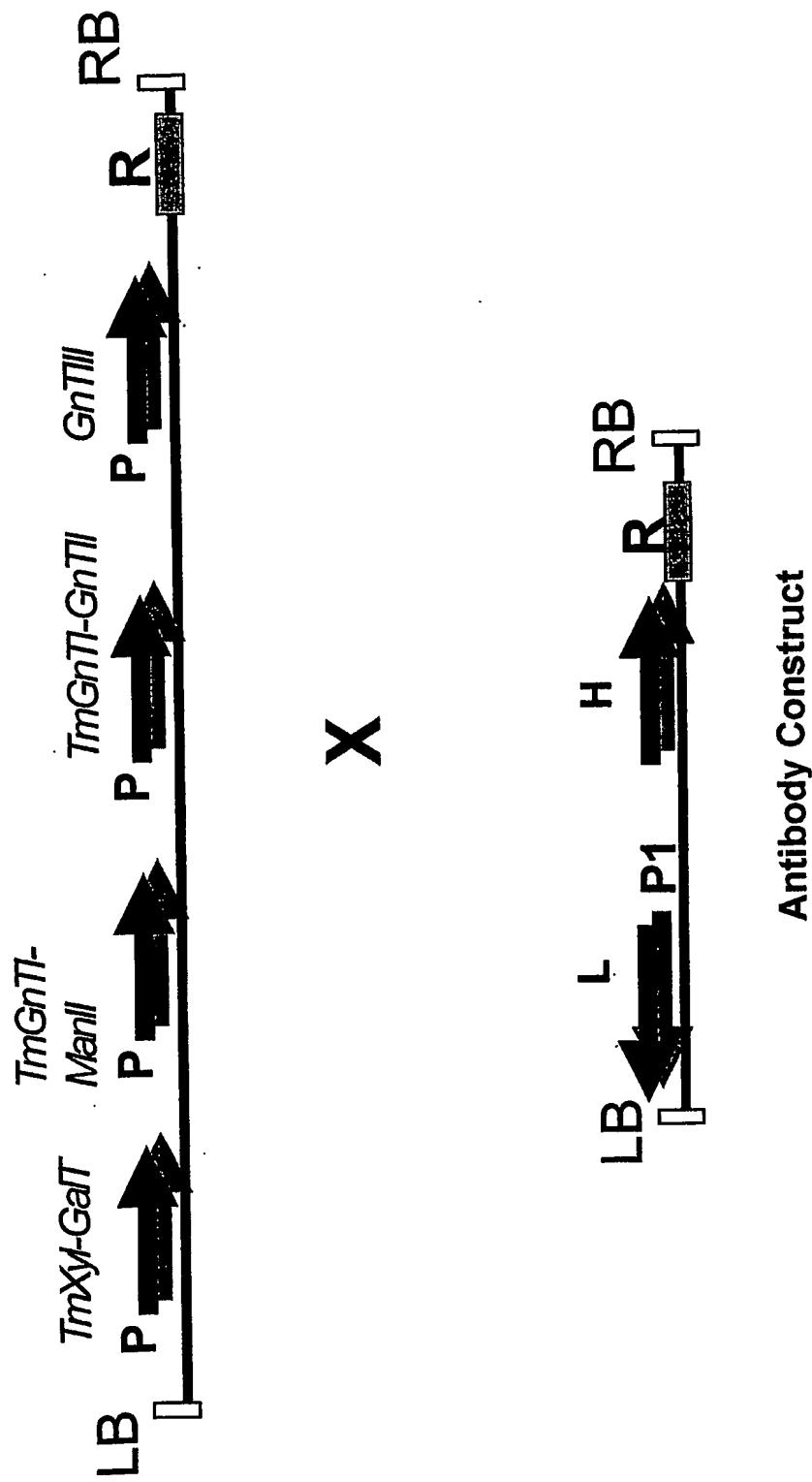


FIG. 2





Antibody Construct

FIG. 4

FIG. 5

6/57

atgaggcttcggagccgtcctgagccgcacgcgcgcgtatgccaggcgctccctac
M R L R E P L L S G S A A M P G A S L Q
cgggcctgcgcctgctcgccgtctgcgcctgcacacctggcgtcacccctcgttac
R A C R L L V A V C A L H L G V T L V Y
tacctggctggccgcacactgagccgcctgccccaaactggtcggagtcctccacaccgctg
Y L A G R D L S R L P Q L V G V S T P L
cagggccgctcaacagtgcgcgcgcattggcagtcctccggggagctccggaccgga
Q G G S N S A A A I G Q S S G E L R T G
ggggccggccgcgcctcctctaggcgcctcccccagccgcgcgggtggcgaactcc
G A R P P P L G A S S Q P R P G G D S
agccccagtcgtggattctggccctggccccgctagcaacttgacctcggtcccaactgccc
S P V V D S G P G P A S N L T S V P V P
cacaccaccgcactgtcgctggccgcctgcccctgaggagtccccgctgctgtggccccc
H T T A L S L P A C P E E S P L L V G P
atgctgatttagttaacatgcctgtggacactggagtcgtggcaaaaggcagaacccaaat
M L I E F N M P V D L E L V A K Q N P N
gtgaagatggcgccgctatgccccaggactgcgtctccctcacaagggtggccatc
V K M G G R Y A P R D C V S P H K V A I
atcattccatcccaaccggcaggagcacctaactgacttgcttatatttgccacc
I I P F R N R Q E H L K Y W L Y Y L H P
gtcctgcagccacgcagactggactatggcatctatgttataaccaggcggagacact
V L Q R Q Q L D Y G I Y V I N Q A G D T
atattcaatcgctcaagactcctcaatgttggcttcaagaaggcctgaaggactatgac
I F N R A K L L N V G F Q E A L K D Y D
tacacctgcttggttttagtgcgtggacacttccaaatgaatgaccataatgcgtac
Y T C F V F S D V D L I P M N D H N A Y
aggtgttttcacagccacggcacattccgttgcataatggataagttggattcagccta
R C F S Q P R H I S V A M D K F G F S L
Ctttatgttcagtatggagggtgtctctgcctcaagtaacaacagttctaaccatc
P Y V Q Y F G G V S A L S K Q Q F L T I
aatggatttcataataatttggggctggggaggagaagatgatgacattttaacaga
N G F P N N Y W G W G G E D D D I F N R
tttagtttttagaggcatgtctatatctcgcccaatgctgtggcggagggtgcgtac
L V F R G M S I S R P N A V V G R C R M
atccgccactcaagagacaagaaaaatgaacccaatcctcagaggttgaccgaattgca
I R H S R D K K N E P N P Q R F D R I A
cacacaaaggagacaatgcctctgtatgggttgaactcactcacctaccagggtcgat
H T K E T M L S D G L N S L T Y Q V L D
gtacagagataccattgtatacccaatcacagtgacatcgggacaccgagacttag
V Q R Y P L Y T Q I T V D I G T P S -

FIG. 6

7/57

MRLREPLLSGAAMPGASLQARCLLVAACALHLGVTLVYLYLAGRDLSSRLPQLVGVSPTPLQGGNSNSAAIIGQSSGELRTGGARPPLGL
ASSOPRPGDSSPVVDSGGPAGASNLTSVPVPHHTTALSIPACPEESPLLVGPMLJEFNMPVLDIELVAKQNPVNKMGGRYAPRDCVSPHKV
AIIIFRNRRQEHLKYWLYY1.HPVLQRQQLDYGIYGIY'VINQAGDTIFNRAKLLNVGFQEALKDYDTCTFVFSDVDLIPMNDHNAVRCCS
QPRHTSVAMDKFGFLSPVQYFGGVSA LSKQQFLTINGFPNNYWGWGGEDDDITNLVFRGMSISRPNAAVGRCRMIRHSRDKKNEPN
PORFDRAHTKETM LSDGLNLSLT YQVLDVQRYPLYTQITVDIGTPS

FIG. 7

88

atgagtaaacggaatccgaagattctgaagattttctgtatatgttacttctcaactct
M S K R N P K I L K I F L Y M L L L N S
ctctttctcatcatctacttcgtttcactcatcgctcgccccggagcagtacag
L F L I I Y F V F H S S S F S P E Q S Q
cctccatataaccacgttcagtgaataaccatcgccatcgcccagtcctccgg
P P H I Y H V S V N N Q S A I G Q S S G
gagctccggaccggaggggccggccgcctccttaggcgcctcccccagccgcgc
E L R T G G A R P P P L G A S S Q P R
ccgggtggcactccagccagtcgtggattctggccctggcccccgttagcaacttgacc
P G G D S S P V V D S G P G P A S N L T
tcggcccagtccccacaccaccgcactgtcgctgccccctgcctgaggagtcggc
S V P V P H T T A L S L P A C P E E S P
CtgcttgtggcccatgtgattgagttAACATGCCTGTGGACCTGGAGCTCGTGGC
L L V G P M L I E F N M P V D L E L V A
Aaggagaaccaaattgtgaagatggccggccctatgccccaggactgcgtctccct
K Q N P N V K M G G R Y A P R D C V S P
cacaagggtggccatcatcattccatccgcacccggcaggagcacctaagtaactggcta
H K V A I I I P F R N R Q E H L K Y W L
tattatttgcacccagtcctgcagcgcaggcagctggactatggcatctatgttatcaac
Y Y L H P V L Q R Q Q L D Y G I Y V I N
caggccggagacactatattcaatcgtgtaagctcctcaatgttgcttcagaagcc
Q A G D T I F N R A K L L N V G F Q E A
ttgaaggactatgactacacctgcttgcattgtgatgcgtggacccattccaaatgaat
L K D Y D Y T C F V F S D V D L I P M N
gaccataatgcgtacagggttttcacagccacggcacattccgttcaatggataag
D H N A Y R C F S Q P R H I S V A M D K
tttggattcagcctaccttatgttcaatgttggaggtctctgtcttaagtaacaa
F G F S L P Y V Q Y F G G V S A L S K Q
cagtttctaaccatcaatggatttctaataattttggggctggggaggagaagatgat
Q F L T I N G F P N N Y W G W G G E D D
gacatttttaacagatttagtttagaggcatgtctatatctcgcccaaattgtgtggtc
D I F N R L V F R G M S I S R P N A V V
gggagggtgtcgcatgtccgcactcaagagacaagaaaaatgaacccaatcctcagagg
G R C R M I R H S R D K K N E P N P Q R
tttgcaccaattgcacacacaaaggagacaatgtctgtatggttgaactcactcacc
F D R I A H T K E T M L S D G L N S L T
taccaggtgctggatgtacagagataccattgtataccaaatcacagtggacatcg
Y Q V L D V Q R Y P L Y T Q I T V D I G
acaccgagctag
T P S -

FIG. 9

MSKRNPKILKJFLYMLLINSLFLIDYFVFHSSFSPEQSQQPHIYHVSVNNQSAIGQSSGELRTGGARPPPLGASSQPRPGGDSSPVVDSG
PGPASNLTSPVPHTAISLPACPEESPLLYGPMLJEFNMPVIDELVAKQNPNVKMGGRYAPRDCVSPHKVAILPFRNRQEHIKYWLY
YLHPVLQRQQLDYGYVINQAGDTIFNRAKLLNVGFQEALKDYDTCFVSDVDLIPMNDHNAYRCFSOPRHISVAMDKGFGFSLPYVQ
YFGGVVSALSKQQFLTNGFPNNYWGWGEDDDIFNRLVFRGMSISRPNAVVGRCRMIRHSRDKKNEPNPQRFDRIAHTKETMLSDGLN
SLTYQVLDVQRYPLYTQITVDIGTPS

FIG. 10

10/57

FIG. 11

ccatggatgagacgcataagctttctatgttctgtatggccggcctgtgcctcata
 M V M R R Y K L F L M F C M A G L C L I
 tccttcctgcacttcttcaagaccctgtcctatgtcaccttccccgagaactggcctcc
 S F L H F F K T L S Y V T F P R E L A S
 ctcagccctaacctggtgtccagcttttctgaaacaatgccccgtcacgccccaggcc
 L S P N L V S S F F W N N A P V T P Q A
 agccccgagccaggaggccctgacctgctgcgtaccccactctactcccactcgccccctg
 S P E P G G P D L L R T P L Y S H S P L
 ctgcagccgtgcccccagcaaggcgccgaggagactccaccgggtggacttggtgctg
 L Q P L P P S K A A E E L H R V D L V L
 cccgaggacaccaccgagtttcgtgcgcaccaaggccggcgtctgcttcaaacc
 P E D T T E Y F V R T K A G G V C F K P
 ggcaccaagatgtggagaggcccccggacggccggaggagaagcctgagggggcc
 G T K M L E R P P P G R P E E K P E G A
 aacggctcctccggccggccaccccgtaacctctgagcggggagcgcacgggg
 N G S S A R R P P R Y L L S A R E R T G
 ggcaggccggccggcgaagtgggtggagtgcgctgtgcctgcccgtggcacggacc
 G R G A R R K W V E C V C L P G W H G P
 agctcgccgtgcccactgtggtgcaagtactccaacctgcccaccaaggagcggctgg
 S C G V P T V V Q Y S N L P T K E R L V
 cccaggagggtgcgcgcgcataacgcaccaacgtcaaccacgagttcgaccc
 P R E V P R R V I N A I N V N H E F D L
 ctggacgtgcgttccacgagctgggacgtggacgcgccttggactcc
 L D V R F H E L G D V V D A F V V C E S
 aacttcacggcttatgggagccggccgctcaagttccggagatgtgaccaatggc
 N F T A Y G E P R P L K F R E M L T N G
 accttcagtagtacatccggcacaagggtgtctatgtcttgcaccacttccggcc
 T F E Y I R H K V L Y V F L D H F P P G
 ggcggcaggacggctggatgcgcgacgactacctgcgcacccctcaccaggacggc
 G R Q D G W I A D D Y L R T F L T Q D G
 gtctcgccgtgcgcacccgtggccgacgactacccctgcgcacccctcaccaggacgg
 V S R L R N L R P D D V F I I D D A D E
 atcccgcccggtgacggcgctttcaagtcatacgatggctggaccggccctc
 I P A R D G V L F L K L Y D G W T E P F
 gcctccacatgcgcacccgtggccgacgactacccctgcgcacccctggag
 A F H M R K S L Y G F F P W K Q P G T L E
 gtgggtcaggctgcacggcatgtgcaggcgtatggctggacggcatccgc
 V V S G C T V D M L Q A V Y G L D G I R
 ctgcgcgcggccactacccatgcgcacccgtggccgacgactacccctgcgcaccc
 L R R R Q Y Y T M P N F R Q Y E N R T G
 cacatccgtgcaggctggcgtggccgacccctgcacttcgcggctggactgctcc
 H I L V Q W S L G S P L H F A G W H C S
 tgggtcggccggccactacccatgcgcacccgtggccgacggcatcc
 W C F T P E G I Y F K L V S A Q N G D F
 ccacgctgggtgactacgaggacaagcgggacccgtggactacccatccgcggctgatccgc
 P R W G D Y E D K R D L N Y I R G L I R
 accggggctgggtcgacggcacgcaggactacccgcctgcagaccggcggac
 T G G W F D G T Q Q E Y P P A D P S E H
 atgtatgcgcacccgtggactacccatgcgcacccgtggccgacccctgcgcac
 M Y A P K Y L L K N Y D R F H Y L L D N
 ccattaccaggagccaggacacggcgccggcgccggatggcc
 P Y Q E P R S T A A G G W R H R G P E G
 aggccggccggccggggcaaaactggacggcggaaatgcgaacaaaactcatctcagaa
 R P P A R G K L D E A E V E Q K L I S E
 gaggatctgaatttaggatcc
 E D L N - D

FIG. 12

MVMMRYKLF MFCMAGLCLJ SFLHFFKTL S YVTFPRELAS LSPNL VSSFF WNNAPVTPQA SPEPGGPDLI RTPLYSHSPL
LQPLPPSKAA EELHRVDSL VL PEDTTEYFVR TKAGGVCFKP GTKMLERPPP GRPEEKPEGA NGSSARRPR YLLSARERTG
GRGARRKWVE CVCLPGWHGP SCGVPTVVQY SNLPTKERL V PREVPRRVTN AINVNTHEFDL LDVRFHEL GD VVDAFVVCES
NFTAYGEPPRPLKFREMLTNG TFEYIRHKVL YVFLDHFPFG GRQDGWIADD YLRNFLTQDG VSRLRNLRPD DVFIUDDADE
IPARDGVFLKL YDGWTEPF AFHMRSKSLYG FFWKQPGTLE VVSGCTVDMQ QAVYGLDGR LRRRQYYTAMP NFRQYENRTG
HILVQWSLGS PLHFAGWHCS WCFTEPEGYT KL VSAQNGDF PRWGDYEDKR DLNYIRGLR TGGWFDTQQ EYPPADPSEH
MYAPKYLKN YDRFHYLLDN PYQEPRSTAA GGWRHRGPEG RPPARGKLDE AEVEQKLISE EDLN

FIG. 13

FIG. 14

catgagtaaacggaatccgaagattctgaagattttctgtatatgttacttctcaactct
 M S K R N P K I L K I F L Y M L L L N S
 ctctttctcatcatctacttcgttttctactcatcgtcgccccactggaggccctgacctg
 L F L I I Y F V F H S S S F S P E Q S Q
 cctcctcatatataccacgttctactgaataaccatccgcacatggaggccctgacctg
 P P H I Y H V S V N N Q S A H G G P D L
 ctgcgtaccccactctactcccactcgccccctgctcagccgctgcccaggcaaggcg
 L R T P L Y S H S P L L Q P L P P S K A
 gccgaggagctccaccggggacttggacttggacttggacttggacttggacttggacttgg
 A E E L H R V D L V L P E D T T E Y F V
 cgcaccaaggccggcggcgtctgttcaaaccggcacaaggatgtggagaggccccc
 R T K A G G V C F K P G T K M L E R P P
 ccggacggccggaggagaaggcctgagggggccaacggctcctggccggccaccc
 P G R P E E K P E G A N G S S A R R P P
 cggtaacctcctgagcggggggagcgcacggggggcaggcggccggcgaatgggg
 R Y L L S A R E R T G G R G A R R K W V
 gagtgctgtgcctggccggacttggacggaccagctggcgtgcccactgtggcag
 E C V C L P G W H G P S C G V P T V V Q
 tactccaacctgcccaccaaggagcggctgtggcggcagggtgcggcggcgtcattc
 Y S N L P T K E R L V P R E V P R R V I
 aacccatcaacgtcaaccacgagttcacctgtggacgtgcgtcgttccacgactggc
 N A I N V N H E F D L L D V R F H E L G
 gacgtggacgccttggacttggacttggacttggacttggacttggacttggacttgg
 D V V D A F V V C E S N F T A Y G E P R
 ccgctcaagttccggagatgtgaccaatggcaccttcgactacatccggcacaagg
 P L K F R E M L T N G T F E Y I R H K V
 ctctatgttccctggaccacttccggccggcggcggcggcggacttggacttggactt
 L Y V F L D H F P P G G R Q D G W I A D
 gactacctgcgcaccccttcacccaggacggcgtctcgccggacttggacttgg
 D Y L R T F L T Q D G V S R L R N L R P
 gacgacgttccatcattgacgtcgacgagatccggccggacttggacttggacttgg
 D D V F I I D D A D E I P A R D G V L F
 Ctcaagctcacatggacttggaccacttccggccggacttggacttggacttggactt
 L K L Y D G W T E P F A F H M R K S L Y
 ggcttcttggaaaggcggggcaccctggagggtgtcaggctgcacggacttggactt
 G F F W K Q P G T L E V V S G C T V D M
 ctgcaggcagtgtatggctggacggcatccgcctgcgcggccggacttggacttgg
 L Q A V Y G L D G I R L R R R Q Y Y T M
 cccaaacttccacatggacttggacttggacttggacttggacttggacttggactt
 P N F R Q Y E N R T G H I L V Q W S L G
 agcccccctgcacttcggcgtggacttggacttggacttggacttggacttggactt
 S P L H F A G W H C S W C F T P E G I Y
 ttcaagctcggtccggcagaatggcacttccacatggacttggacttggacttgg
 F K L V S A Q N G D F P R W G D Y E D K
 cgggacacttgcacatccggcgttgcacttgcacatggacttggacttggacttgg
 R D L N Y I R G L I R T G G W F D G T Q
 caggagttccgcctgcacgggggggggggggggggggggggggggggggggggggg
 Q E Y P P A D P S E H M Y A P K Y L L K
 aactacgaccgttccactacatggacttggacttggacttggacttggacttggact
 N Y D R F H Y L L D N P Y Q E P R S T A
 gccccccctgcacttcggcgtggacttggacttggacttggacttggacttggact
 A G G W R H R G P E G R P P A R G K L D
 gaggcggaaacttgcacaaaactcatctcagaagaggatctgaatttaggatcc
 E A E V E Q K L I S E E D L N - D

FIG. 15

MSKRNPKILK IFLYMLLNS LFLIYFVFH SSSFSPEQSQ PPHIYHVSVN NQSAHGGPDL LRTPLYSHSP LLQPLPPSKA
AEELHRVDLV LPEDTTEYFYRTKAGGYCFK PGTKMILERPP PGRPEEKPEG ANGSSARRPP RYLLSARERT GGRGARRKWW
ECVCLPGWHGPSCGVPTVVQ YSNLPTKERL VPREVPRRVNHEFD LLDVDRFHELG DVVDASFVCE SNFTAYGEPR
PLKFREMLTN GTFEYIRHKV LYVFLDHFPP GGRQDGWIAID DYLRTFLTQD GVSRLRNLRP DVVFIDDAD EIPARDGVLF
LKLYDGWTEP FAFHMRKSLY GFFWKQPGTL EVVSGCTVDM LQAVYGLDCI RLRRRQYYTM PNFRQYENRT GHLYQWSLG
SPLHFAGWHCSWCFTPEGIY FKL VSAQNGDFPRWGDYEDK RDLNRYRGLIRTGGWFDTQ QEYPPADPSE HMYAPKYLK
NYDRFHILLDNPYQEPRSTA AGGWRHRGPE GRPPARGKLD EAEEVEQKLISEEDLN

FIG. 16

GGC GGCGCTGAGGCATCGCAGATCTAACCAATTACGATAACGCTTGGGTACACTGATTITGTTCA
 TGGTTACATATCTTGTTATATGCTATCTTAAGGATCTGCACAAAGATTATTGTTGATGTTGATGGGG
 CTCAGAAGATTIGATACTGATACTCTTAAGGAGATACCCAGGCCAGGATTATATTCAAGACAATCAA
 TTACACGTGGTCAAACTCGTTATCTTCAAGGATGAGCCAGAATCTTATAGAAATGATGCAATCGAGA
 ATGTTGGCGCATGGCTTGTGGCTCAATTCTACATATCACACAAAGAATCGACCGTATTGTACCCCTCTT
 CCATAAGGAAAACACAAATTGCAAGATCGCTTTCACATCGCAGTAACATAAGGTATTCAAAAATGGCTAA
 GAAGTTGGATAACAAATTGACAACACTTCCATTCTGTTATAAAATTCAACACACAAAGCCCCGTAATCAA
 GAGTCGTGCCCATGTACCGAAATTAACTCTTATTGGTATTGGTATTGGCTTAAGGCTCAAGAGTACGTGGGGTACC
 ACATATAGGAAGGTAACAAATTACTGCAAGATAAGGATAACGCTTACATCCCTTACCCACGGAAAGGATAAGA
 TATAAGGACCCACCCCTGCCACGTTGTCACATCGTCATGGTGGTTAATGATAAGGGATTACATCCCTCATGTTGG
 ACATGATGCAATTGTCATGAGGCCACAGGATCCAATGGCCACAGGAACAGTAAGAATTGAGATAATTGATIT
 GTCCGGTTAGAATTAGCAAAACATTATAAAAGGTGTTGATCAATAGGAACACTAACATTCACTCATGGATTCA
 CCATTCTCTTAAGTATCTAGAAACCATGGCGAGGATCTCGTGTGACTTGAGATTCTCATCCGGCAGCCTT
 CATGTTCACTCCAGATGAGGCTTTCAGACGCCAATCACAGTATGCAAGTACAGATCAAAACAGTCGGGATTGTTGCC
 TCTGAGAACCATGGCACTAGTCAAATTGCGAGGCCTCATAGATGAAAGTTAGCATCAATCCAGAGTTG
 TCGAAGGATAATGAAAGAACCGCCAGGACGAAAGAACCTGGCAGCTTAAGGATCTAATCCAGAGTTGAAAAAAAAGG
 AATAGCAAAACTCACTCAAGGTGGAGCCATGGATTCCAATTCAAGGCGCCGCTGCGTTGATAATCACAAACTAAAGATCTA
 TACGATAGGATTGAGTTCTGATAACAGATGGTGGCCATGGAAACCAAGGGTTGGAGAGTTACGTAATAAGACGATG
 AGTGGGAGAAAGAGACAAATTCTCGTGGTCCATTCTCAATAACGGATCCCTCATTAAGGATCTGAGAAAGTT
 GGAGTATTCTAGAGACAATTCCAGACATAATTCTGACACCAATTCTGTTAATGAGCTTAAATGCTTAAATTC
 TTGCTTTGGTTTAATATTAACTCTCCATTGGTAAACCGTGAACAAACCTTAAGGATCTGTTTAAATGCTTAAATTC
 GACGGTCAATTAAACTCTATAACCAAACCTCTTGTGGGTTCTGTTTAAAGTTCGTTGATGAAACAGAGTTCT
 AGAAGTTCGTTCTTGGAAAAATTGAAAGTCTTGGAGCTAAAGTTGTTTATTACTGGGTTTGAGATTGA
 AGGATAGCTAGAAATCTTATTTGTTGAATATGTTAATAGGATTCAAGAAAGTTTATAATG
 GGAGGAGATGTCATCTGGAGAGATGGAGAGTCACCTAACAAACAAGACGCTTCACTAACAAACAAGGTT

FIG. 17

AAGGATGGGCAGCTAGAGATTGGGAGGTGGCTGGGTATGAATGATGAGGCTAAATTCAACATTTCACATTGCCATAA
 TTGAACAGATAAGCAGAGGGTAATATGTGGCTGAATGACACAATTGGGTATTCCCTAAGAACATTCTGGCTATAGA
 TCCCTTGGCTATTCAACCATTGGCTTATCTCTCCGGGTATGGTTTGAAAACATGCTTATTCAAAAGGACT
 CATTACCGAGCTCAAGAAAGACCTTGCCCCAGCATTAAGAACATTCTGAATAATTGGCGTTCAGAGCTGGATGCTATGG
 AAACACACAGATACTTGTTCATATGATGCCGTTTATTCAACAGATACTCCACACACTTGTGGACCAAGCCTGC
 AATTGCTGTCACTTGTGATTTCGGCTGGGATGGGGATTIAAGTATGAACCTTGTCCATGGGAAAGCACCCAGTG
 GAGACCACACTAGAAAATGTGCAGGAGGGCATTAAAGGCTTCTGGATCAATACAGGAAAAAATCCACTCTAAATC
 GAACATAACACTCTTACCCCTGGAGATGATTAGGTACATTAGTATCGATGAAGGCCAGGGCTCAGTTCCG
 TAACCTACAGATGTTGTTGATCACATCAACTCTAACCTCTGGAGATGATTAGGTACATTAGTATCGATGA
 GATTATTTCAGAACAGTCCGAGAAGCAGACAGAGTGAATTATTCTCGTCCTGGTGGAGGTGGCTCGGTCAAGG
 TTGTTGGTTCCCTCTCTGTCAAGGTGACCTCTTACATATGCAAGATAAGGACAACAAGACTATTGGAGTGGTTATTAA
 TGTTCAAGAACCTTCTCTAACAGCTGTTGATCGTGTGCTGAGCATACCCCTCGTGGAGCTGAGATCATGATGTC
 TTCTGGCTAGGGTATTGCCAATTCAATGTAAGAAATTCCAACAAGTTACGTATAAGTGTGACTGCTGCAA
 GAAGAAAATCTGGCTCTTCCAGCACCATTGAGGGTAACCTGGAAACTGCTAAGGATATTGTGGTACAAGATTACGG
 CACCCGGATGCATACTCTCATTGCAAGAACCTTCAGATCTTATGTCATAAGCAATTGAAAGTTCTTCITGGGATCCGC
 CACGAGAAAGAAAATCTGATCAATCCCATCATTTCAGGGCAGAGCAATTGAGATCAAAAGTATGATGCTGGC
 CAGTTCAAGCCAATTGCTGCCGGAAAGGAAATTGGCACACAGTTATACTCTTCATCCCATCAGAACAGAGAG
 AGAGGGGGTGGTGAAGGGTTGTTGTTAACCGGGCTGAAATCTCGGTTGGACTCAAACGGACTTGTGCTCCTAGC
 CAAATTCTCCCTGAAGTGCAGCATGACGGATACCAAACATTGCTGAGAACATATTGCTTACATTGCTTACATTG
 TCCAGGCTCTGGTCTGAGTTGACCCATTTCCTCTTCATATTGCTTACATTGCTTACATTGCTTACATTGCTTAC
 CAAATACGCTTCTGAGTTGACCCATTTCCTCTTCATATTGCTTACATTGCTTACATTGCTTACATTGCTTAC
 GAGATCCGAAATGAACATCAGACTCTTGTGTTGATGTGAAGAACGGATCAACTGCGGAAGATAAGTCCATAGAAACG
 GATCAGAGACTGTGTTGGAGAAGAGATAAGGTATGACTCTAGTCTGAGGAGCTGGAGCTTACCTGTTCAACCCAGA
 TGGTGAAGGCTAGCCAATTGTTCAACCTGATGGACATGTAGTCACCTCTGAGGGCTGCTGGTTCAAGGAAGTCCTC
 TCTTACCCCTAAACCAAATCACCCTCTCTCAGAAAACCTGCTTACACTGGAGGTAAATACGGCTTC
 AGGATCAAGTGGTCTGAGAGATAAGAATATCATGTTGAGCTTGGTAATGATTTGATGACCCGGAAATTGATTGTC
 GTACAAGACTGATGTTGACAACAAGGGCTCTTCAATGTTCCATTCAAGATCTCAATGTTCCAAATGAGGAGAGAAACT

FIG. 17 Cont.

TATGATAAGATCCCTTCAAGGAAACTACTACCCAATGGCCATCTCGCATTTATCCAAGGATCCAATGGTCAGA
GATTCTCCGTCAATCTCGTCACTCGTCAATCTCGTCAAAAGGGGTTGGAGATTATGCTGGACAG
ACGGTTGGTTCTCGTGAAGGGTCTAGGGCAAGGGTGTGATGGATAACCGGCAATGACCGTGGTATTTCAC
CTCTTGCAGAACCCCTGCTCAAGGAGACCCCTGCTTCAAAACACATTCATTGCCAAGAACATCTGCGTGT
ACCTATAGGGTGGCTCACTTAACACTACCCATAAACACATTCATTGCCAAGAACATCTGCGTGT
TCCACAATACGGGTCCTTGCCTTIAGCCAACCGTTACCATGTGACCTCCACATTGTAATTTCAAGGTTCT
CGTCCATCCAAATACTCTCAGCAATTGGAAAGACAAGCCAAGGGTCCATCTCGATCTCTCAATAGACGGCTGGG
ATTCAAGCTTATGGCCATAAGGAAGACAAGTAACCTGCAACAAGCATGGCTAATGAACCAGTAACCTTCCGACAT
GTCAAAAGATCTTGCAAGCTTCAAAAGGTAACCTCAACTGAAATCTCTTGCAAGGAAGATATGGAGATTCTGGG
TACGATGACCAAGAGCTACCTCGAGATAGTTCAACAGCCACGGGAAGGACGTGTCTCGATCTCTCAATGGAAATAC
GAGCCTATAAGCTTGAACACTGGACCTCACAAGTGAACCTGCTGAAGAATTCGCTAGAGTCCGCTAGAGTCCGCAA
TCTCTACAAATCTATCTCTCTTCTATTTCAGAAATAATGTTGAGTAGTTCCAGATAAGGGAAATTAGGGT
TCTTATAGGGTTCGCTCATGTGTTGAGCATATAAGAAACCCCTTAAGTATGTATTGTATTGTAAATACTCTT
CAATAAAAATTCATAACCAAACCCGGCCCTCGAGGGGATCGCAGATCTCATTACCGTTAGA
AGCATAGTTAAAATCTAAAGCTTGTCCCTTAATTCTAGTCATTIACATTGTTGGGTCTACATTATAATGAATT
TCTAATGCAAAATACAGAAATTCAAATGTAACACATGTAAACATACGTTATCTCCGCTCTG
TGTGTTGTTAACTTGAAGTTATCATAAAGTATCATAAAGAACCACAAATACACTAGTAAATCTATGAGAACGGCAG
AAACAAGAGATCTAAGATTTCATTGTAAGTATAGGAATAATAATCTCTTATCTGATTAAATGAAATCCACATG
TCACACTCTCATTTGTCACAAAGATCACAACCTTATCTCAATATTCACAAACTTGTATATTCAC
TCTTTCACTAGCCCCACAAAATACTTGTCCCCTTATTTGCCACCCCTTGTATTAAATTATTCTGTGGAGCT
AAGTGTTCATATTATTCTCCTCAAAAACAAAAACAAAAAGAGAAACCATTGGCGAGGAACCTCGT
GTGACTTGAGATTCTCTCACTCCGGCAGCTTCATGTTCATCTCACATCCAGATGAGGCTTTCCAGACGCCAATC
ACAGATGCAAGATCGCCCTCAGRTCCGGCTATCGAAACTGCAACTAGTCAAATGCGAGGGCTCATAGAT
GAAGGTAGCATCAAAACAGTGGGGATTGTTGCCCTCGAAGATAIGAAGAACCCGCCAGGAAGAAACITGGCAGC
TTAAGGGATCTAATCCAGACGTGTTGAAAAGGAATAGCAAACACTCAAGGTTGGAGCCATGGCTCTAAGGTT
GCATAGAAGGAACCAATTTCGGCTTACGAAATACGGATCTGTTCCGGATTGGCAAAAGATCGTGTGGTTATCGTC
TGTATGTGCATAATCGGGCTCAGTATTTCGACTTGAAGAGATGAATTAGGATTGTTAGTATGCTGAAAGGTAA
CATTTGAGGAGTATTGTTAGTCATGTTAGTGTGTTAGTAAAGGTATAAAGGTGTTAGTGTGAAAGGAGTTAAAGT
CATTTGAGGAGTATTGTTAGTCATGTTAGTGTGTTAGTAAAGGTGTTAGTGTGAAAGGAGTTAAAGT

FIG. 17 Cont.

GAAACAGATTCTGCCCTATCGCTCATATATACTAGCTTCCGGGTGTGACCTGAAATGATTGTAAG
AACAAAGGGTGAATGAGGCCAAAGGGCAATTGTGAAGGTAATTCCTGATCATGGAAATCATCGGTCTCCGAAGAATTG
TATCTTGAAGCATTCACTGGTGGGATGATGAACACTGTAATGGGATGGGGTTGGAAAGAGACTAAAGGACATGAGGG
GCATATCCTTTCATTGAAGAAGATCATTTCTGTTCCCTAATCGTAACATACAGACTCTACGAGGGCTG
AAACCCGCAAAAGTGTCTGACTGTTCTGCTGCTAATTAGCACCGTCTGATGTGAAGTCAAGAGGAGAAGGGCTTG
AAAGTTGGTGCAGAGAGAAATGGGAAATGTTGGGTATTCCTTAATAGAAGTGTGGGAGAATATTCATTCAGAA
GGCAAGAGAGTTGGTTCTTGATGATTACAACCTGGGATAAAACGATGTTGGCAACGGGTTTCCCCGTCGGTTGGT
TCCCCGGTGTACACATTCGAGGGCCTAGGACTAGTGGGGTACACTTGGAAAATGTGGGTTGCATCAAGGTTAGAG
GAGATGAGGGTGAATTGCAATCGATAATGGGTCGTAACATAGAAGTTAAGGAAACAGATAAAAGTTGGAAACATAAA
AGAAGGATGGGGAGTTCGGGTGTTAAAGCATCAAGCGGGTTAAAGCGGGTTTCAAGGGTGGGGTT
GATGATAAGGGACCGACATTATGTTGGATTTGCCACTATGTTACAGCGTAGGCACTGCATCTCCATGAA
ACGGATCCGGCTAGAGTCCGCAA AAAATCACCAAGTCTACAAATCTAATCTCTCTATTTTCTCAGAAATAA
TGTGTGAGTAGTTGCTTATAGGGGAAATTAGGGAAAGATAAGGGAAATCTAATCCCTAATCCTAAATTCTAATCCTAA
CTCTTAATTAA

FIG. 17 Cont.

GGCGCGCCTCGAGGGGATCGCAGATCTAACCAATTACGATAACGCTTGGGTACACTTGATTGTTGTTTCAG
TGGTACATATCTTGTTTATATGCTATCTTAAGGATCTGCACAAAGATTATTGTGTGATGTTCTGTGATGGGG
CTCAGAAGATTGTGATATGATACTCTAACATCTTGTGAGATACCAGCCAGGATTATTCAGTAAGACAATCAAAT
TTACGTGTTCAAACTCGTTATCTTCATTCAAGGGATGAGCCAGAATCTTATAGAATGATTGCAATCGAGAAT
ATGTCGGCCGATATGCCCTTCAATATTCTACATATCACACAAGAATCGACCGTATTGTAACCTCTT
CCATAAGGAAAACACAATATGAGATTCAGTACATGCCACATGCACTAACATAGGTATTCAAAATGGCTAAAA
GAAGGTTGGATAACAAATTGACAACATTCCATTTCATTCTGTTATATAAAATTCAACACACACAAAGGCCCGTAATCAA
GAGTCTGCCCATGTAAGGAAAATAACTCTTATTGGTATTGGCTTAAGGCCAGCTCAGAGTACGTGGGGTAC
ACATAAGGAAGGTAACAAAATACTGCAAGATAGCCCCATAACGTTACCGCCTCTCCTTACCGAAAGAGATAAGA
TATAAGACCCACCCCTGCCACGTTCACATCGTCATGGTGGTTAAATGATAAGGGATTAACATCCCTCTATGTTGTGG
ACATGATGCAATGTAATGTCATGAGCCACAGGATCCAATGGCCACAGGAACAGTAAGAATGTAGATAAGTATTGATTT
GTCCGGTTAGATAGCAAAACATTAAAGGTGTTGATCAATAGGAACACTAATTCACTATAGGATTCATAGAAGT
CCATTCCTCCCTAAAGTATCTAGAAACCATTGGCGAGGGATCTCGTGTGACTTGAGATTCTCTCATCCGGCAGCCTT
CATGTTCACTACATCCAGATGAGGCTTTCAGCAGCAATCACAGTATGCAAGATGCGCTCATAGTAAACAGTGCCT
TCTGAGAACCATTCGCACTAGTCAAATCGAGGCTCATAGATGAAGGTTAGCATAAACAGTGCAGGATTGTTGCC
TCGAAGAGATGAAAGAACCGGCCAGGACGAAGAACACTGTGCAAGCTTAAGGATCTAACTCCAGACGTTGAAAGAAGG
AATAGCAAAACTCACTCAAGGTTGGAGCCATGGATTCCAATTCAGGCGCCGTCGTTGATATACAACAAAGATCTA
TACGATAGGATTGAGTTCTGATACAGATGGTGGCCATGGAAAGGTTGAGTTACGTTAAAGACGATG
AGTGGGAGAAAGAGAAAGCTCAAAATCTTCGTTGTCCTCATTCICATAAACGATCCCTGGTGAATTGACTGTAGA
GGAGTATTACAGAGACAATCAGACATTCAGACATTGACACCATTGTTGAGACTTATCTAAAGGTATGACGAAAGGTT
TTGCTTTGGTTTAATTTAATTCTCCCATGGTTATCCCGTGAACAACTTAAATGTCCTAAATTCAT
GACGTGCAATTAAACTCTATAACCAAACCTCTTGTGGGGTTGTTGAGCTAAAGTTGTTGATGAAAGAGTTCT
AGAAGTTCGTCTTGGAAATTGTTGAGCTAAAGTTGTTGAAATATGTTTAATAGGATTCAGAAGAAAGTTATATG
AGGATAGCTAGAATCTTATTGTTGAGCTTAAAGGAGATGGTGGAGACGCTTACCTAATCTGGAGAGATGGTGG
GGGAGGAGATGTCATACCTAACAGAAGGCTTACCTAATACAGAAGGAGATGGTGGAGACGCTTACCTAACAGAAGG
AAGGATGGGAGCTAGGTTATGAAATGTTGAGGCTTAATTCCACATTATGTCAGGCTTAATTCCACATTATGCCCATAA

FIG. 18

TTGAAACAGATAGCAGAGGGTAATAATGTGGCTGAATGACACAATTGGGGTTATTCCCTAAGAATTCTGGCTATAGA
 TCCCTTGGCTATTCAACCATTCACTTCATGCCATTATCTCTCCGGGTATGGCTTATGGGTTTGAAAAACATGCTTATTCAAAGGACT
 CATTACCGAGCTCAAGAAAGACCTTGCCAGCATAGAAATCTTGAATATTTCAGATATGCCATTACAGCTGGATGCTATGG
 AAACACACAGATATCTTGTGATGCCGTATTACGATACACACTTGGGACCCAGCTGC
 AATTGCTGTCAGTTGATTTCGCTCGGATGCCGGATTAAAGCTTGTGATCAATACAGGAAAAAAATCCAACTCTATATC
 GAGACCAACTAGAAAATGTGCAGGGAGAGGGCATTAAAGCTTCTGGGATGATTAGGTACATTAGTATCGATGAGCCAGGCTCAG
 GAACATAACACTCTTACCTCTGGAGATGATTAGGTACATTAGTATCGATGAGCCAGGCTCAG
 TAACCTACAGATGTTGATCACATCAACTCTAAACGCTAAACCTAGTCTAAACGCGAGAACAGAAAGTTGGTACTTTGGAG
 GATTATTCAAGAACAGTCCGAGAACAGAGTGAATTCTCGTCCTGGTGGGGTTGGCTGAGGTGAGGTTGGCTCAG
 TTTGGTTTCCCTTCTGTCAAGGTGACTCTTACATATGCCAGATAGGCAACAAAGACTATTGGAGTTGTTATA
 TGTTCAAGAACCTTCTCAAGGTGTTGAGGAAATTCCAACAAAGTTTACGTATAAGTTGA
 TTCTGCTAGGTTATTGCCATCGAAATTCAATGTGAGAAATTCTCGTGGCTGAGGTTACGGTACATTAGGATTTACGTATAAGTTGA
 GAAGAAATCTGGCTCTTCCAGACCCATGATGGGGTAACCTGGAAACTGCTTAAGGATTATGTTGATCAAGATTACGG
 CACCCGGATGCGATACTCTATTGCAAGAACCTTCAGATCTGATCTGCAACTTTCGAGGGCAGAGCAAATGAGATCAAAAGTATG
 CACGAGAAAGAAAATCTGATCAATCCCCATCATTTTCGAGGGCAGAGCAAATGAGATCAAAAGTATG
 CAGTTCACAAGCCAATTGCTGCCGGGAAGGAAATTGCGCACACAGTTATCTCAATCCCATTCAGAACAGACCGAG
 AGAGGAGGTGGTGAACGGTTGTTGITAACCGCGCTGAAATCTGGGTTTGGACTCAAAACTGGGACTTGTGTCCTCTAGC
 CAAATTCCTGAAAGTGCAGCATGACGATACCAAACACTTACCGGCAGACATCGCCTTATCTGGAAAGCTTCA
 TCCCAGCTCTGGCTCTGAGAACATAATTCAATTGCTTAATGGGAATTGTCGAGTGTGAGAAAGCTAATCTGGCTCTAAACT
 CAAATACGCTCTGGCTCTGAGTTGACCCATTCCCTCCATATTCCTGCTCCAAACTGGACAACGACGTTACT
 GAGATCCGAAATGAAACATCAGACTCTTGTGTTGATGTGAAAGAACGGATCACTGGGGAAAGATAGTCCCATAGAAACG
 GATCAGAGACTGTTGGGGAGAAAGAGATAAGGTATGTACTCTAGTCCAGAGAGTGGAGCTTACCTGTTCA
 TGGTGAAGGCTCAGCCAAATTGTTCAACCTGATGGACATGTTAGTCACCTGAGGGTCTGCTGGGTCAGAACAG
 TCTTACCCCTAAACCAAATGGGAGAAATCACCCTCTCTCAGAAAACACTCGTCTTACACTGGGGTAATACGCTTC
 AGGATCAAGTGGTCAGAGATAAGAATACTATGTTGAGCTTCTGGTAATGATTTGATGACCGGGAAATTGTTGTCAGA
 GTACAAGAAGACTGATGTTGACAACAGAAAGGTCTCTTATTCAATGTTGTTCAATGTTGAGGAGAGAAACT
 TATGATAAGATCCCTCTTCAAGGAAACTACCAATGCCATCTCGCATTATCCAAGGATCCAATGGTCAGA

FIG. 18 Cont.

GATTCTCCGTGCACTCTCGTCAATCTCGGTGTGCAAGGCCTCAAAGAGGGTTGGAGATTATGCTGGACAG
ACGGTTGGTTCTGATGACGGACGGGGCTAGGGCAAGGTGTGATGGATAACCGGCCAATGACCGTGGTATTTCAC
CTTCCTTGCGGAATCTAACATTCTCAAGCAGACCCCTGCTTCCAACACTAAACCGAGGAACCCCTCGCTCTCTC
ACCTCATAGGTGCTCACTTAAACTACCCATAAACACATTCATGGCCAAGAACATATCTGTGCGGT
TCCACAATACGGTTCCTTGCTCCATTAGCCAACCGITACCATGTGACCTCCACATGTAAATTTCAAAGGGTCC
CGTCCCATCCAAATACTCTCAGCAATTGGAAAGAACAGCAAGGCTTCTTATCCCTCAAAATAGACCGAGCTGGG
ATTCAAGCTTATGCCATAAAGGAAGACAAGTAACAGCACAAAGCATGGCTAATGAACCAGTAACCTTTCGGACAT
GTTCAAAGATCTTGCAGCTTCAAAGGTAAAACCAACTTCACTGAAATCTCTTGCAGAAAGATAAGGAGATTCTGGG
TACGATGACCAAGAGCTACCTCGAGATAAGTTCACAGCCACGGGAAGGACGGTGTCTCGATCTCTCCCATTGGAAATAC
GAGGCTTATAAGCTTGAACCTGCAACAGTGAACCTGCTGAAGATCCGCTAGAGTCCGAAAATCACCAAGTC
TCTCTTACAAATCTATCTCTCTATTTCAGAATAATGTGTGAGTAGTCCCAGATAAGGGAAATTAGGGT
TCITATAGGGTTCTGCTCATGTGTTGAGCATAAGAAACCCCTAGTATGTAAATTCTAT
CAATAAAATTCTAAATCCTAAACCAAAATCCGGAGAGACCCCTCTTAATAA

FIG. 18 Cont.

CCATGGGAGGATCTCGTGTGACTTGTACAGATTCTCATCCGGCAGCTTCTCATCTACATCCAGATGAG
 GCTTTTCAGACGGCAATTCACAGATGCAAGTATGGCAGATCGGCTCAGTTCCGCTATCGAATCTGAGAACCATTCAGTCAA
 ATGGCAGGGCCTCATAGATGAAGCTTAGCATCAAAACAGTCCGGGATTTGTCGCCCTCGAAGATATGAAAGAACGGCCAGG
 ACGAAGAACCTGTGTCAGGCTTAAGGATCTAACAGACGTTTGAAATTCACAAACTAAAGATCTTACAGATAGGATTGAGTTCTGAT
 AGCCATGGGATTCAAATTCAGGGCCGGTCTGGTGAATTCACAAACTAAAGATCTTACAGATAGGATTGAGTTCTGAT
 ACAGATGGTGGCCATGGAAACAAGGTGGAGAGTTACGGTATAAAGACGATGAGTTGAGGGATATTACAGAGACAATCCAG
 TCTTCGTTGTTUCTCATCTCATAACGATCCTGGTGGAAATTGACTGTAGAGGATTTACAGAGACAATTTAA
 ACATATTCTGACACCATTGTGAGACTTATCTAACGGTATGACGAAAGTTTGCTTTGGTTTAATTTAA
 TTCTCTCCCATGGTTATCCGTTAACAAATCTTAATGTCTTAAAATCTCATGACGTCATTAAACTCTATAACCAA
 ACTTCCTTGCIGGGGTTCTGTTTTAGTTTCGTTGATGAAAACAGAGTTCTAGAAAGTTCGTTCTTTGGAAAAATT
 TGAAGTCTTGGAGCTAAAGTTGTTTATTACTGGGTTTGAGGATTGAAAGGATAGCTAGAATCTTATTTGTT
 TGGGGTTTGGTTGAATAATGGGTTAACAGAAAGTTTATATGGGAGGAGATCTGATCATCTGGAGAG
 ATGGTGGAGAGACGCTCACCTAATAAACAAAGCTTACATGAGGGCTAATTGAATGAGGGTTTGAATTAAGGATGGGTTAATA
 GGAGGGTGGCTGGGTATTGAATGAGGGCTAACATTAATTCACATTAATTGCCATAATTGAAACAGATAGGAGGGTAAATA
 TGTGGCTGAATGACACAAATTGGGTTATTCCCTAACAGAACATTTCAGGCTATAGATCCCTTGGCTATTCAAAACCAT
 GGCTTATCTCTCCGGCGTATGGGTTTGGAAAACATGCTTATTCAAAAGGACTCATTACGGAGCTCAAGAACGACCTT
 GCCCAGCATAAGAATCTTGAATAATTGGGTCAAGAGCTGGGATGCTATGGAAACCACAGATATCTTGTTCATA
 TGATGCCGTTTATTCAACGATATCCCACACACTTGTGGACCCAGGAGGACCACTAGCTGCAATTGCTGTAGTTGATC
 TCGGATGCGGGATTAAAGTATGAACCTTGTCCATGGGAAAGCACCTGCAATTGCTGTTGATC
 GAGAGGGCATTAAGCTTCTGGATCAATAACAGGAAAAAAATCCACTCTATACTGAAACTAATACACTCTTAACTCT
 TTGGGAGATGATTTAGGTACATTAAGTATCGATGAAGGCCAGGCTCAGTTCCGTAACCTACCAAGATGTTGTTGATCA
 CATCAACTCTAACCTAGTCTAACGGAGAACAAAGTTGGTACTTTGGAGGATTATTCAAAACAGTCCGGAGAA
 GAAGCAGACAGAGTGAATTATCTCGTCTGGTGGGCTCTGGTCAAGGGTTGGTTTATTATGTTGAGTGGTTTAAAGC
 GTGACTCTTACATATGCAGATAGGCAACAAGACTATTGGAGTGGTTTAAAGGAGCTGAGATCATGATGTCATTTCGCTAGGTTATTGCCATCGA
 ATTCAATGTTGAGAAATTCCAACAGTTTACGTATAAGGTTGACTGCTGCAAGAACAAATCTGGCTCTTTCCAGC

FIG. 19

ACCATGATGGGTAACCTGGAACTGCTAAGGATTATGTGGTACAAGGATTACGGCACCCGGATGCATACTTCATTGCA
 AGACCCCTCAGATCTTTATGTCTAAAGCAATCGAAAGTCTTCTTGGATTCGCCACGAGAAAATCTGATCAA
 TCCCCATCATTTCGAGGGCAGAGCAATGAGATCAAAGTATGAGATCAAAGTATGAGATCAAAGCAGCTCAGTGGCT
 GGGAAAGGAATTTCGACACAGTTAACCTCITCAATCCATCAGAACAGAACAGAGGAGGGTGGTGAACGGTTGTTGT
 TAACCGGGCTGAAATCTCGGTTTGGACTCAAACCTGGAACTTCCAGCTTGGAAAGCTTCATCCCAAGCTCTGGTCT
 GACGATAACCAAACTATCACCAGGACATCGCCCTTACTGGAAAGCTTCATCCAGCTCTGGTCTGGTCTGGAGAACAT
 ATTCAATTGCTAATGGGAATGTGCGAGTGTGAGAAAGCTTACCGTCTAAACTCAAATACGCTCTGAGTTTGACCC
 ATTCCITGTCTCCTCATATTCTGCTCCAAACTGGACAAACGACGTTACTGAGATCCGAAATGAAACATCAGACT
 CTTGGTGTGATGTGAAGAACGGATCACTGGGAAGATAGTCCAGAGAGTGGAGCTTACCTGTTCAAAACCAAGATGGGAG
 AGATAGGTATGTACTCTAGTCCAGAGAGTGGGTCTGCTGGTCTGCTGGTCAAGAAGTCTCTTACCCCTAAACCAAAATGGGAG
 ACCTGATGGACATGTAGTCACCTCTGAGGGTCTGAGGGATCACTGGAGGTTAATACGGCTTCAAGGATCAAGTGGTCTGAGATAAGAAT
 AAATCACCCCTCTCAGAAAACCTCGTCTTACACTGGAGGTTAATGAGGAAATGGTCCGGTACAAGAAGACTGATGTTGAGAACAA
 ATCATGGTGAACCTCTTGGTAATGATTGATGACCCGGGAATTGATGAGCAGGGAGAGAAACTTATGATAAGATCCCTCTCAAGGA
 GAAGGGTCTTCIATTCAATGGTCTCAATCTCAGATCTCAATGGTCTCCAAATGAGCAGGGAGAGAAACTTATGAGGTTCAAGGATCCA
 AACTACTACCCAAATGCCCACATCTCGCATITATCCAAAGGATCCAATGGTCAAGAGATTCTCGTCAACTCTCGTCAAT
 CTCTCGGTGTGCAAGGCTCAAAAGGGTTGGATTGCTGGACAGACGGFTGGTCTGTAAGGCTTCTCGCTTCTCTCAACCTTAAC
 GGGTCTAGGGCAAGGTGTGATGGATAAACCGCGCAATGACCCGTTACCTCTCACCTCATAGGTGCTCACCTTAAC
 CAAGCAGACCCCTGCTTCCAACACTAACCCGAGGAACCCCTCGCTTCTCTCACCTCATAGGTGCTCACCTTAAC
 ACCCCATAAAACACATTCAATTGCCAAGAACCGCAAGAACGAAACCTTCAAGGTTCCACATGTTGACCTCCATCGTCC
 TTAGGCCAAACCGTTACCATGTCGCTCTTATCTCAATAGACGAGCTTGGGATTGCCATTGCTTAAAGGAA
 TTGGAAAGAACACAAGCCAAAGGGTTCGCTCTTACGAGCTTGGGATTGCCATTGCTTAAAGGAA
 GACAAGTAAACTGCAACAGCATGGCTAATGAAACCGTAAACTTCCGACATGTTCAAAGATCTTGCAAGCTTC
 GGTAAGAACCAACTCACTGAATCTTGCAAGAAGATATGGAGGTTACGATGACCAAGAGGCTACCTCGA
 GATAGTTCACAGCCACGGGAAGGACGTTGCTCGATCTCCCATTGAAATACGAGCTTAAAGCTTGAACCTGCGAC
 CTCACAAAGTGAACCTGCTGTAAGAT

FIG. 19 Cont.

FIG. 20

CCATGGCGAGGATCTCGTGTGACTTCAGCTTCATCCCCAGCTTCATGTCATGTCATCTACATCCAGATGAG
GCTTTCCAGACGCATCACAGTATGCAGATTCAGATCTCGCCTCAAGTCCGCTATCGAAATCTGAGAACCATTCGACTAGTC
ATGCCAGGGCCTCATAGATGAAGGTTAGCATCAAACAGTCGGGATTGTTGCCCTCGAAGATACTGAGAACCATTCGACTAGTC
ACGAAAGAACTGTCAGCTTAAGGATCTAATCCAGACGTTGAAAAAAGGAATAGCAAAGACTCAACTCAAGGTGG
AGCCATGGCTCAAGGTTGCATAGAAGGAACCATTTTCGCCTAGAAATAACGGATCTGTCGGATTGGCAAAA
GATCGTGTGGTTATCGTCTTGTATGTGCATAATCGGGCTCAGTATTTCGAGTCACAGTGGAAAGGTTTCGAGG
TAAAGGTATAAGTGAGACATTGTTGATGTTGATGGTACCTTGAAAGAGATGAATAAGGATTGGAGAG
TATTAAGTTCGTCAAGTGAACAGATACTGGCTTATCGCCTTATCGTACTAGCTCCGGGTGTG
ACCTGAAATGATTGTAAGAACAGGGTGTGAGGGCAAAAGGGGCAATTGTGAAGGTAATCCTGATCAGTATGGGAATC
ATCGGTCTCCGAAGATGTATCTTGAAGGCATCACTGGTGGTGGATGATGAACACTGTATGGGATGGGTGGAAAGA
GACTAAAGGACATGAGGGCATATCCRTTTCATGAAGAAGATCATTTCTGTTCCAATGCCCATACTGTAACATA
CAGACTCTTACGAGGCTGAAACCCGCAAAGTGTCCCTGACTGTTTGTGCTGCTAATTAGCACCCGCTGTGAAAGT
CAAGAGGAGAAGGGCTGAAAGGTTGGTGCAGAGAGAATGGGAATATGGAAATAGTGTGGTAATTCTTTAATAGAAGTGTGTG
GGAGAATATTCACTCAGAAGGCAAAGAGAGTTTGTGATTACAACCTGGGATATAACCGATGTGGCAACG
GTTTCCCCGTCGTTGGTCCCCGGTGTACACATGCGAGGGCCTAGGACTAGTGCGGTACACCTTGGAAAATGTG
GGTGCATCAAGGTAGAGGAGATGAGGGTGTGATTGCATCGGATAATGGGTGTAAACATAGAAAGTTAAGGAAACAGA
TAAAGTTGTGAACATAAGAAGGATGGGGAGTGGGGAGTTCGGGTGTATAAGCATCAAGCGGGTTATAAAGGCCGGTTTCGAA
GGTGGGAGGTTGGGGGATGATAGGGACATTATGTTGGGATTTGCCACATGTTACGTATGTC
GCAGTGGCATCTCCATGAAACGGATCC

FIG. 21

CCATGGCAGGGATCTCGTGTGACTTGTGAGATTCTCTCATCTCACATCCAGATGAG
GCTTTCCAGACGCCAATCACAGTATGCAGATCCGCTCAGTCTGAGAACCATTCAGTCAA
ATGGGAGGCCCTCATAGATGAAAGTTAGCCATCAAACAGTCGGGATTGTTGCCCTCGGAAGATATGAAAGAACCGCCAGG
ACGAAGAACCTTGTGCACTTAAGGATCTAACAGACGTTTGAAGAAAAGGAAATAGCAAACACTCAAGGTGG
AGCCATGG

FIG. 22A

CCATGGCGAGGGAGCAGATCAGTGGTAGCAGCAGGCAAATGGAGGTACTGCAACCCCTCCTATTACTTGAA
GCGCCAAAGCGTCTGCTCTGCATCGTTTCGTTCTGGGACCGTCAAACCTCTC
GTCAAGAGCACCGTTGAAAATTCTGAGCTGCAGAAAAGAAGTGACTGATTGTTAA
ATAACAAACAAGGTGTAACCTGGAAAAGTGAACCATGG

FIG. 22B

GGCGGCCCTCGAGGCGATCGCAGATAACAAAATTGAAATCGCACAGATCGATCTCTTGGAGATTCTAT
 ACCTAGAAATGGAGACGATTTCAAATCTCTGTAAAAATTCTGGTTTCTTGACGGAAAGAACGACGACTCC
 AATAATTTCGGTTAGTACTGAACCGGAAAGTTGACTGGTGCACCAATTAAATTGATGTCAGCCATTAAACTA
 GGATTITGTATTCATGGCCTTATCTGTGAGCCATTAAATTGATGTCAGCCCTAAACTAAATCCGAACGGTTA
 TTTCAGCGATCCGGACGGTTGTATTAGCATAATAGCATCAATTATGTAGCAGTGGTGAATCTCG1CAAACCAAG
 TAAAGCTAGATCTGGACCCGTGAATTGGTGCACAGAAAGCACATGGTGTGATAATTTCACCGGTACGATTAGAAAC
 TTGAGAAACACATTGATAATCCGATATAAAATCCGTTACCATCGTGCCTATAAAATTAA
 TATCAATAGCGTACACCGGTGAAAGACTGACAATTATTATCTTTGAAATTGGAGCTCAAGTTGAAATTGGAG
 AAGCTAGAGAGTTCTGATAACCATTGGCGAGAGGGAGAGATCAGTGGTAGCAGCAGCAATTGGAGGTACT
 GCAACCCCTCTCTTACTCTGAAAGGCCAAAGCGTCTTGTCTCTCATCGTITTCGTTTGTCTCTTGC
 TTCTGGACCGTCAAACCTCTCGTCAGAGGACCCAGGGTGAAAATTCTGAGCTGCAAGAAAGTAAGTGA
 AAAAATTGGTGGATGTTAAATAACAAACAAAGGTGTTACCTGGGAAACTGACTTGGGGACCATGGGACAGA
 TGCCCTGGCTGCTGTAGTGGTTATGGCCTGCAAGACTCTTGTGAGACTATCTGAAAGGACTGTAAATCGTTAAC
 ATATCAAACCTCCCGTGTCTCAAATAATCCTCTATTATATCTCAGGGATGGATCTGATCAAGCTGTCAAGAGCAAG
 TCATTGAGCTATAATCAAATTACATATGCAACACTTGGGATTGAAACCAGTGGTCACTGAAAGGCCTGGCAAC
 TGACTGCGTACTACAAGATTGACGTCAGTCACTACAAGTGGCACTGGGACTCTGGCACTGGTTTACAACAA
 AGTGATTATACTAGAAAGATGATAATGGAAATTGCTCCAGACTCTTGTGATTACTTGGGCTGCAAGCTG
 GATAGGGATAAAACCAATTATGGCTGCTCATCATGGAAATGATAATTGGACAGAAGGCAAGTGGCTG
 CGCTATAACCGATCAGATTITCTGGCTTGGGATGGTCAAGAGATCGACTGGGATGAGTTACCAAC
 GTGGCCAAAGGCTTACTGGGATGAIITGGCTGACATAAGGAAACCTAAAGGGCGCCAATTTCAGTC
 GTCTGTAGAACATACAAATTGGTGAACATGGGTCTAGTTGGGACAGTTTCAGTCAGTATCTGGAAACCTATAA
 AGCTAAACCGATGTCAGGGTGAAGCAAGGACCTGGGATAACCTGACAGAGGGAAACTATACCAAGTACTT
 TTCTGGCTTAGTGGAGACAAGCACGACCAATTCAAGGTTCTGACCTTGTCTAAAGGCTCAAAACATAAGGATGAT
 GTTCGTTACCGGTATAAGACCAAGTAGAGTTGAACCGCATGGGAAATTGGTATATTGAAAGAAATGGAAGG
 ATGGTGTGCCCTCGAACAGCATATAAGGAGTAGTGGTGTTCGAATCCAGACAACAAAGACGGTGTATTCTGGTGG
 GCCAGATTCTGTAAATGCAAGCTGGAAATTGCGGATCCGCTAGAGTCCGCAAAATCACCAGTCT
 CTCTCTACAAATCTATCTCTCTATTTCAGAATAATGTTGAGTAGTGTGAGTGTGAAATTAGGGTT

FIG. 23

CTTATAGGGTTTCGCTCATGTGTTGACATAAGAAACCCCTTAGTATGTATTGTAAATTGTAAATACTCTATC
AATAAAATTCTAACTCTAAACCAAATCCCGCGCTCGAGGGATCTCAATTACGATA
GCTTGGGTACACTGATTGTTCAAGTGGTACATATATCTTGTGTTATGCTATCTTAAGGATCTGCACA
AAGATTATTGTTGATGGTTCTTGATGGGCTCAGAAGATTGATATGATAACACTCTAATCTTAGGAGATAACCAGC
CAGGATTATTCAAGACAATCAAAATTACGTGTCAAACTCGTTATCTTCAATTCAAGGATGAGCCAGA
ATCTTATAGAAATGATGCAATCGAGAATATGTTGGCGATAATGCCCTTGTGGCTCAATACTACATATCAC
ACAAGGAATGCGACCCGTAATTGTAACCCCTTCCATAAAGGAAAACACAATATGCAAGATGCTTTTCCCACATGAGT
AACATAATTAGGTTATCAAAATGGCTAAAGAAGTTGGATAACAAATTGACAACATTTCCTGTTATAAA
TTTCAACACACAAAGCCCGTAATCAAGAGGTCTGCCCATGTAAGAAATAACTCTTATTATTGGTATTGGCCT
AAGCCCAAGCTCAGAGTACGTGGGGTACACATATAGGAAGGTAAACAAATACTGCAAGATAAGCCCCATAACGTAC
CAGCCTCTCCATTACCAAGAGATAAGATAAGACCCACCCCTGCCACAGTGTACATCGTCATGGGGTTAATGTA
TAAGGGATTACATCCCTCTATGTTGGACATGATGCAATGTAATGTCATGAGCCACAGGATCCAATGGCCACAGG
AACGTAAGAAATGTAGATAGATTGATTGATTGATTGATTCTGTCAGTAGAAACACATTAAAGGTGTATCAATAGGA
ACTAATTCACTCATGGATTCAAGAAGTCCATTCCCTTAAGTACTGCAACCCTCCATTACTGTAAGGCTCTGC
GTGGGTAGCAGCAGCAAAATGGAGGTACTGCAACCCTCCATTCTGGGACCCGTCAAACCTCGTCAGAGGCCACAGGTGA
TCTTCATCGTTTTCGTTGTCTCTTCGTTGAGGTTACGTATAAGACGATGAGTGGTACCTCT
TCTGAGCTGAGAAAGTGAATTGAAAAATTGGTGTGATGATTGAAAAACTAACAAACAGGTGGTACCT
GGGAAAAGTGAACGGGACCATGGATTCCAATTCAAGGTTGGTCCATGGAAACAAAGGTTGGAGAGTTACGTATAAGACGATGAGTGGG
GGATGAGTTCTGATAACAGATGGTGTGTTGAGGTTACGTATAAGACGATGAGTGGTACCT
GAAAGAGAAAGCTCAAAATCTCGITGTCCTCATTAACGATCCTGGTGGAAAATTGACTGTAGAGGTAT
TATCAGAGACAATTCAAGACATTCTTGACACCATGGTTATCCCGTGTGAAACAATTCTAAATGTCTTAAATTGTTAGITTCGTTGATGAAAGTCTGA
TGGTTTAATTCTAAACTCTAAACAAACTCTTGTGGGTTCTGTTTGTGAAATTGAGGTTTGTGAGATTGAAAGGATAG
TAAACTCTATAACCAAACCTCTTGTGGGTTCTGTTTGTGAAATTGAGGTTTGTGAGATTGAAAGGATAG
CGTTCTTGTGAAAAATTGAAGTCATTGGAGCTAAAGTGTGTTTGTGAGATTGAAAGGATAG
CTAGAATCTTAACTCTTGTGGGGTTCTGTTTGTGAAATTGAGGTTTGTGAGATTGAAAGGAG
ATGTCATATCTGGAGAGATGGTGGAGAGACGCTTCACTTAATAACAAAGAAGCTTGTGACTAAATTGITAAGGATG
GGCAGCTAGAGATTGTTGGAGGTGGCTGGGGTTATGATGAGGCTAATTCCACATTATTGCCCCATAATTGAAACA

FIG. 23 Cont.

GATAGCAGAGGGTAATATGTGGCTGAATGACACAATTGGGGTTATTCCATAAGAATTCTGGCTATAGATCCCCTT
 GGCTTATTCAACCATTGGCTTATCTTCTCCGGCTATGGGTTTGTGAAAACATGCTTATTCAAAGGACTCATTACG
 AGCTCAAGAAAGACCTTGCCAGCATAGAATCTTGAAATATTGGCGTCAAGGGCTGGGATGCTATGGAAACAC
 AGATATCTTGTTCATATGATGCCGTTATTACGATATCCCACACACTTGGGACCAAGGCCTGCAATTG
 TGTCAAGTTGATTTCGCTCGGATGCCGATTAAGTAGTGAACCTTGTCCATTGGGGAAAGCACCCAGTGGAGACCA
 CACTAGAAAATGTGCAAGGAGGGCATTAAGGTTCTGGGATGATTAGTACATTAGTATCGATGAAGGGAAAGC
 TACACCTCTTATACCTCTTGGGATGATTAGTACATTAGTATCGATGAAGGGCTCAAGTCCGTAAC
 CAGATGTTGTTGATCACATCAACTCTAATCCTAGTCTAAACGGCAGAAAGCAAAAGTTGGTACCTTGGAGGATTATT
 TCAGAACAGTCCGGAGAACAGAGTGAATTATTCGTCCTGGTGAAGGTGGCTCAGGTGG
 TTCCCTCTCTGTCAAGGTGACTCTTACATATGAGATAGGCAACAAAGACTATTGGAGTTATTATGTTCA
 AGACCTTCTCAAAAGCTGTGATCGTGTGACTACCCCTCGAGCATACCCCTCGTGGAGCTGAGATCATGATGT
 TAGTTTATTGCCATTCGAATTCAATGTGAGAAATTCCAACAAAGTTACGTATAAGTTGACTGCTGCAAGAAGAAA
 TCTGGCTCTTCCAGACCATGATGGGGTAACTGGAAACTGCTAAAGGATTATGGGTACAAGGATTACGGCACCCGG
 ATGCAACTTCATGCAAAGACCCTCAGATCTTATGTCATAAGCAATCGAAAGTCTTCTGGATCCGGCACGAGA
 AAGAAAAATCTGATCAATCCCCATCATTTTCGAGGCAGAGCAAATGAGATCAAAGTATGAGAT
 CAAGCCAATTGCTGCCGGGAAGGAAATTTCGACACAGTTATACTCTCAATCCATGAGAACAGACGGAGAGGGAG
 GTGGTGACGGGTTGTTGAAACCGCGCTGAAATCTCGGTTACTGGACTCAAACCTGGACTTCTAGCCAAATT
 CTCCCTGAAGTGCAGGATGACGATAACCAAACATTACGCCAGACATGCCCTTACTGGAAAGGCTCCATCCAGC
 TCTGGTCTGAGAACATAATTCAATTGCTAATGGGAATGTGCGAGTGTGAGAAAGCTACTCCGTTAAACTCAAAATAC
 GCTTCTGAGTTGACCCATTCTGTCTCCATATTCCCTGTCACCTGCTGGAAAGATAGTCCATTAGAAACCGATGGTGAA
 GAAAATGAAACATCAGACTCTGTGTTGATGTGAAGAACGGATCACTGCGGAAAGTACCTGTCACCTGTT
 GACTGTTGTGGAGAAAGAGATAAGGTATGTACTCTAGTCCAGAGAGTGGAGCTTACCTGTTCAAACCAACT
 GCTCAGCCAATTGTTCAACCTGATGGACATGTAGTCAAGAAGTCTTCTTACCTGCTGGGCTGCTGGTCAAGAAGTCTTCT
 CTAACCAAAATGGGAGAAATCACCCCTCTCAGAAAACCTCGTCTTACACTGGAGGTAATACGGCTTCAGGATCA
 AGTGGTCGAGATAAGAAATCATGTTGATGTTGAAATTGATTGATGACCCGGAAATTGATTGTCGGGTACAAG
 ACTGATGTTGACAACAGAAAGGTCTTCAATTCAATGAGATCTCAATTGTTCCAATGAGCAGGAGAGAAACTTATGATA
 AGATCCCTCAAGGAAACTACCAATGCCATCTCGCATTTATCCAAGGATCCAATGTCAGGATTCT

FIG. 23 Cont.

CGTGCACCTCGTCAATCTCTGGTGTGCAAGCCTCAAAGAGGGTTGGAGATTATGCTGGACAGACGGTTG
 GTTCGGTGTGATGACGGACGGGGTCTAGGGCAAGGGTGTGATGGATAACC CGC AATGACCCGGCAACCTCTTG
 CGGAATCTAAACATTTCTCAAGCAGACCCCTGCTTCCAACACTAACCCCTGCTTCCACCTCTCCTCACCTCAT
 AGGTGCTCACTAAACCTAACCCATAAACACATTCATTGCCAAGAAAGACATACTGGTGCCTGTTCCACAA
 TACGGGTTCCCTTGCTCCCTTAGCCAACCGTAAACCATGTAACCTGCTCCACATTGTAACATTCAAGGGTCTCGCTCC
 CCAAAATACCTCTCAGCAATTGGAAAGACAAGGCCAAGGGTTCGCTCTTATCCTCAATAGACGAGCTGGGATTCAAGC
 TTATGCCATAAGGAAGACAAGTAAACTGCACAAAGCATTGGCTAATGAAACCGATAACCTTCCGACATGTTC
 GATCTGGCAGGTTCAAAAGGTAAAAACCAACTTCAC TGAAATCTGCAAGAAAGATAATGGAGATCTGGTACGAT
 ACCAAGAGCTACCTCGAGATAAGTTCACAGCCACCGGGAAAGGACGTTGCTCGATCTCTCCATGGAAATACGAGCTTA
 TAAGCTTGAAACTGGCACCTCACAAAGTGAACCTGCTGAAGGATCCGGTAGAGTCCGGTAGAGTCCGCAAAAATACCCAGTCTCT
 ACAAAATCTATCTCTCTCTTATTCCTCCAGAATATGTTGAGTAGTTCCCAAGATAAGGGAAATTAGGGTTCTATA
 GGGTTTCGCTCATGTTGAGCATATAAGGAAACCCCTTAGTTGTTGTTGTTGAAATACCTCTTATCAATAAA
 ATTCTTAATCCCTAAACCAAATCCCGGGCGCCCTGAGGGCGATCGCAGATCTCATTATACCGTTAGAAGCATTAG
 TIAAAATCTAAAGGTTGTCGTTAATTCTAAGTCATTTACATTGTTGGGTCTACATTATTAAATGAAATTCTAATG
 CAAATACAGAATTAAATCTAAACATGTAACATGTTGAATTATGCTAAACATGTAACATGTTGTTGTTGTTG
 TATTAACTTGAAAGTTATCATAAAGAACCCAAATACACTAGTAAATCTATGAGAAGGGCAGGTGGCAACACAAACAAAG
 AGTATCTAAGGATTTCATTGTTGACTATAGGAATAATATCTCTTATCTGATTAAATGAAATCCACATGTTC
 CTCATTGTCACAAAGATCACAACTTCTTCAATATTCAACACTGTTATACCTCCACCAAACTTC
 ACTTAGCCCCACAAATACTTGTCCCCCTTATTCGCCCCCTTATTCGAGGCTAAGTGT
 CATATTATCTCTCTCCTAAACAAAAACAAAAAAAGAGAAGAAACCATGGCGAGAGGGAGGAGATCAGT
 GGGTAGCAGCAGCAAATGGAGGTACTGCAACCCCTTCTTATACCTGAAAGGCCAAAGGGTCTGCT
 TTCATCGTTTCGTTTGTGTTCTTCGTTTCTGGGACCCGTCAAACTCTCGTCAGAGAGCACCAGGGTGAATT
 CTGAGCTGCAAGAAAGAAGTGA CTTGACTGATTGAAATTGGGGATGATTAAATAACAAACAAAGGGGGTACCTCTGG
 GAAAACATGACTGACCTGGGACCATGGCTCTAAGGGTGCATAGAAGGAACCAATTTCGCTCTAGAAATACGGGATCTGTT
 CGGAAATTGGCAAAAGATCGTGTGGITATCGTCTGTTGACTGCAAAATCGGGCTCAAGTATTTCGAGTCACAGTGG
 AAAGTTTGTCGAAAGGTAAAGGTATAAGTGAGACATTGTTGATTGTCACTGATGGTTACCTTGAAGAGATGAA
 TAGGATTGTTGAGAGTATTAAAGTTTGTCAAGTAAACAGATTTCCTCGCCCTTATTCGCT

FIG. 23 Cont.

AGCTTCCGGGTGACCCCTGAATGATTGTAAGAACACAAGGGTGTAGGGCAAAGGGCATTGTGAAGGTAATCCTG
ATCAGTATGGGAATCATCGGTCTCCGAAGATTGTATCTTGAAGCATCACTGGTGGATGATGAACACTGTATG
GGATGGGTGGAAAGGACTAAAGGACATGAGGGGCATACTTTCATTTGAAAGAAGATCATTTCTGTTTCCTAAAT
GCCTATCGTAACATACAGACTCTAACGAGGTGAAACCCGCAAAGTGTCCCTGACTGTTTGCTGCTAATTTAGCAC
CGCTCTGATGTGAAAGTTCAGAGAGAAATGGGAAAGTGTGGGTATTCTTTCTTGTGATTACAACCTGGGATATA
TAATAGAAGTGTGGGAGAATATTCACTAGAAGGGCAAAAGAGAGTTTGTGTTCTTGTGATTACAACCTGGGATATA
ACGATGTGGGCAACGGTTTCCCGTCGTTGGTTCCCGGTGTACACATTGCGAGGGCCTAGGACTAGTGCCTAC
ACTTGGAAAAATGTGGGTGCACTAACGGTAGAGGAGATGAGGGTGTGATTGCAATCGATAATGGGGTGTAAACATAGA
AGTTAAGGAAACAGATAAAGTTGTGAACATAAAAGAAGGATGGGAGTTGGGTGATAAGCATAAGCATTCAAGGGTTAT
AAAGCCGGTTTCGAAGGGTGGGGAGGTGGGGCGATGATAAGGACCCGACATTATGTTGGATTGGCCACTATGT
ATCGTTACAGCAGTAGCAGTCATCTCCATGAAACGGATCCGCTAGAGTCGCCAAAAATCACCAAGTCCTCTAC
AAATCTATCTCTCTATTTCTCAGAATAATGTTGTGAGTAGTTCCAGATAAGGAAATTAGGGTTCTATAGG
GTTTCGCTCATGTGTGAGGATAAACCCCTTAGTATGTATTGTAAATACCTCTATCAATAAAAT
TTCTAATCCTAAACAAATCCGGAGAGACTCTTAATTAA

FIG. 23 Cont.

GGCGGCCCTCGAGGGCATCGCAGATATAACAAATTGAAATCGCACAGATCGATCTCTTGGAGATTCTAT
 ACCTAGAAAATGGAGACGATTTCAAATCTGTAAAAAATCTGGTTCTCTGCAACCCAAATTAAATGTACCGTAACGGACACTCC
 AATATTTCGGTTAGTACTGAACCGGAAGTTGACTGGTCAACCCAAATTAAATGTACCGTAACGGACCAATC
 GGAATTGTGATTCATGGGCCCTTAATCTGTGAGGCCATTAAATTGATGTGACGGCCTAACAATTAAATCCGAACCGGTTA
 TTCAAGCGATCCGGGACGGTTGATTCAAGCCAAATAGCAATCAATTATGTAGCAGTGGTGATCCTCGTCAAAACCAAG
 TAAAGCTAGATCTGGACCGTTGAATTGGGTGCAAGAAAGCACATGTTGIGATAATTAAACCCGTACGATTAGAAAAC
 TTGAGAAACACATTGATAATCGATAAAAACGTCGGCATATAAAATCCGCTTACCATCGTTGCCTATAAATTAA
 TATCAATAGGCCGTACACCGCTGAAGACTGACAATTATATCTTTCGAATTCCGGAGCTCAAGTTGAAATTCGGAG
 AAGCTAGAGGAGTTCTGATAACCATGGCGAGAGGGAGCAGATCGTGGTAGCAGCAGCAAATGGAGGTACT
 GCAACCCCTCCTTAACTACTTGAAGCGCTCCTGCTCTGCTCTGCTCATCGTTTCAGTGCAGAAAGAAGTGA
 TTCTCGGACCCGTCAAACTCTCGTCAGAGAGCACCAGGTGAAATTCTGAGCTGCAGAAAGGGACCATGGGACAGA
 AAAAATTGGTGGATGATTAAATAACAAACAAAGTTGTTACCTCTGGAAAAGACTGACTTGGGACACTATCTGAAAGGACTGT
 TGCTGTGGCTGCTGTAGTGGTTATGGCCTGCAAGTGTGCAAGACTATCTGAAAGGACTGTAAATCAGTTAAC
 ATATCAAACCTCCCGTTGCTTCAAATAATCCTCTTATTIAATATCTCAGGATGGATCTGATCAAGCTGTCAAGAGCAAG
 TCATTGAGCTATAATCAAAATAACATAATGAGGACTTGGGATTITGAACCAAGTGGTCACTGGAAGGCTTGGCGAAC
 TGACTTGGTACTACAAGAATTGCACTGCACTACAAGTGGGACTGGACCAGTGGTCACTTACAACAAATTAGTCG
 AGTGATTAACTAGAAAGATGATAATGGAAATTGCTCCAGACTCTTGTGATTACCTTGGCTGAGCTAGTCTCAIG
 GATAGGGATAAAAACCATATGGCTCTCATGGAAATGATAATGGACAGAAGGAGTGCATGATCCCTATG
 CGCTTACCGATCAGATTTCCTGGCCTTGGGTGGATGCTCAAGAGATGCTCAAGAGGATGAGTTACCAAA
 GTGGCCAAGGTTACTGGGATGATGGCTGAGACTAAGGAAAACCATAAAGGCCCAATTCTGACCCGAA
 GTCTGTAGAACATACAAATTGGTGAACATGGGCTAGTTGGGACAGTTGGGATCTGCACTTCTGGAAACCTATA
 AGCTAAACGATGTGACGGTTGACTGGAAAGCAACGGACCTGGGATAACCTGACAGAGGGAAACTATAACCAAGTACT
 TTCTGGCTTAGTGTGAGACAAGCACGACCAATTCAAGGTTCTGACCTTGGCTTAAAGGCTCAAAACATAAAGGATGAT
 GTTCGTATCCGGTATAAGGCCAAGTAGAGTTGAAACGCAATTGCAAGGGGAATTGGTATATTGAAAGAATGGAAAG
 ATGGTGTGCCTCGAACAGCATAAAGGAGTAGTGGGTGTTCCGAATCCAGACAACAGACCTGGTATTCTGGTGG
 GCCAGATTCTGTAATGCAAGCTGGAAATTCAAGGTTCTGATGCGGATCCTGGCTAGAGTCCGCAAAAATACCAAGTCT
 CTCCTACAAATCTATCTCTCTTCCAGAAATAATGTGTGAGTGTCCCAGATAAGGGAAATTAGGGTT
 CTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAACCCCTAGTGTATTGTAATTGAAAGAATACTCTATC

FIG. 24

GGATCCGATAACAATTGAATCGCACAGATCGATCCTTGGAGATTCACCTAGAAAATGGAGACGATT
TTCAAAATCTGTAAAAATTCTGGTTCTTGACGGAAAGAACGGACTCCAATATTCCGGTTAGTACTGAA
CGGGAAAGTTGACTGTGCAACCAATTAAATGTAACCGTAACGGCACCAATCGGATTTCGGTATTCAATGGGCC
TTATCTGTGAGCCCCATTAAATTGATGTGACCGCCCTAAACTAAATCCGAACGGTTTATTCAAGCGATCCGGACGGTT
TGTATTCAGCCAATAGCAATTATGTAGCAGTGGTGAATCCTCGTCAAACCAGTAAGCTAGATCTGGACCGTT
GAATTGGTGCAGAACAGCACATGTGTGATATTACCCGTACGATTAGAAAACCTGAGAAACACATTGATAATC
GATAAAACCGTCCGATCATATAAATCCGCTTACCATCGTTGCCATAATTAAATTCAGGGAGCTAGAGAGTTCTGATA
ACCATGG

FIG. 25

CCATGGCGAGAGGGAGCAGATCAGTGGGTAGCAGCAGCAAAATGGGAGGTACTGCAACCCCTCCTATTACTTGAA
GCGCCCAAAGCGGTCTTGTCTGCCTTCATCGTTTCGTTGAGCTGAGCTGAGCTGAGAAAGAAGTGA
GTCAGAGGACCAGGTGAAATTCTGAGCTGAGCTGAGAAACTGACTTGGGGACCATGGGACAGATGGCCTGTTGCTAGTG
ATAACAAACAAAGGTGGTACCTCTGGGAAATCTGAAAGGACTGTAAATCAAGTAAACATATCAAACCTCCCGTGC
TATGGCTTGAGTCAGACTGAGCTATCTGAGATGGATCTGATCAAGCTGTCATTGAGCTATAATCAATTAA
AAATATCCCTCATTTATATCTCAGGATGGATCTGATCAAGCTGTCATTGAGCTAAAGGAAACTGACTGCGTACTACAAGATG
CATATAGCAGCACTGGATTITGAACCAAGTGGTCACTGAAAGGCCCTGGGGAACACTGACTGCGTACTACAAGATG
ACGTCACTACAAGTGGCACTGGACCAGTTTACAAACACAAATTAGTCGAGTGATTAACTAGAAGAGATGAT
ATGGAAATTGCTCCAGAACATCTTGTATTACCTTGAGGCTGCAAGCTAGTCTCATGGATAAGGGATAAAACCAATTATGG
CTGCTTCATCATGGAAATGATAATTGGACAGAAGCAGTTTGTGCATGATCCCTATGCCCTATACCGATCAGATT
TCCTGGCTTGGGGATGCTCAAGAGATCGACTTGGGATGAGTTACCCAAGGTGGCCA
GATTGGCTGAGACTAAAGGAAACCAATTCAAGGCCCAATTCAATTGACCCGAAGTCTGAGAACATACAATT
CTGCTTCATCATGGAAACCTAAAGCTAAAGCTAAACGATGTCAGTGGAAACCTTAAGCTAAAGCTAAAG
GATTGGCTGAGACTAAAGGAAACCAATTCAAGGCCCAATTCAATTGACCCGAAGTCTGAGAACATACAATT
GTGAACATGGGTCTAGTTGGGACAGTTTCAGTCAGTGGACAGAGGGATAACCAAGTACTTTCTGGCTTAGTGAGACAAGCA
CTGGAAAGCAAGGACCTGGGATACCTGAGAGGGAAACTACCAAGTACTTTCTGGCTTAGTGAGACAAGCA
CGACCAATTCAAGGTTCTGACCTTGTCTAAAGGCTCAAACATAAGGATGATGTTGCTATCCGGTATAAAGACC
AAGTAGAGTTGAACCGCATTGCAAGGGGAATTGGTATATTGAAAGAATGGAAGGATGGTGGCTCGAACAGCATA
TAAAGGAGTAGTGGTGTATTCCAGACAAAGACCTGATTCTGTAATGCGAGCT
GGAATTTCGAAATTCCCTGATGCGGATCC

FIG. 26

GGCGGGCTCGAGGGGATCGCAGATCTAACCAATTACGATAACGCTTGGGTACACTTGATTITGTTTCAG
 TGGTTACATATATCTTGTATTATGCTATCTTAAGGATCTGCACAAAGATTATITGTTGATCTGATGGGG
 CTCAGAAGATTGATAATGATAACTCTAATCTTAACTCGTTATCTTCAATTCAAAAGGATTATTCAGTAAGACAATCAAAT
 TTACAGTGTGTTCAAACACTCGTTATCTTCAATTCAAAAGGATGAGCCAGAATCTTATAGAATGATGCAATCGAGAAAT
 ATGTTGGGCCGATATGCCCTTGTGGCTTCATATTCTACATATCACACACAAGAATCGACCGTATTGTACCCCTCTT
 CCATAAAGGAAAACACAATTATGCCAGATGCTTTCACATGAGTAACTTCAAAACACAACACACAAAGGCTAAATGCTAA
 GAAGTTGGATAACAAATTGACAACATTCTGTTATAATAAATTTCACAAACACAACACAAAGGCGGTAATCTCA
 GAGTCIGCCCCATGTACCGAAATAACTCTTATTGTGTTATTGGTATTGGCTTAAGGCCAGGCTCAGAGTAGCTGGGGTACC
 ACATATTAGGAAGGTAACAAAATACIGCAAGATAACCCCATAACGTAACCGACCCCTCTTACACGAAGAGATAAGA
 TATAAGACCCACCTGCCACGTTGTCACATCGTCATGGTGGTTAATGATAAGGGATTACATCCCTCATGTTGTTG
 ACATGATGCACTGTAATGTCATGAGCCACAGGATCCAATGGCCACAGGAACGTAAGAATGTAGATAAGATTGATTT
 GTCCGGTTAGATAGCAAACACATTAAAAGGTTGTGTATCAATAGGAACACTAAATTCACTCATGGATTCTAGAAAGT
 CCATTCTCTCTTAAGTATCTAGAAACCATGGCGAGGGAGGAGATCAGTGGTAGCAGCAGGCAATGGAGGTA
 CTGCAACCCCTCCATTACTTGAAGGGCCCAAAGCGGTCTTGCCTCTGCTCTTCATCGTTTGTGTCCTCTTC
 GTTTCTGGACCCGTCAAACCTCGTCAGAGGCCACCGAGGTGAAATTCTGAGCTGCAAGAAAGAAGTGA
 TGAAAATTGGTGGATGTTAAATAACAAAGGTTGTTACCTCTGGAAAACACTGACCTGGGACCATGGATTC
 CAATTAGGGCCCGTCGTGTTGATATCACAAACTAAAGATCTACGATAAGGATTGAGTTGAGTGGAGAAAGAGAAAGCT
 CCATGGAAACAAAGGTTGGAGAGTTACGTTAAAGGACGATGAGTTACGTTAAAGGACGTTACGTTAAAGGAGCT
 CTCATCTCATAACGATCCTGGTGGAAATTGACTGTAGAGGAGTATTATCAGAGACAATCCAGACATATTCTGA
 CACCATGTTGAGACTTATCTAAGGTAAGCTTGGTTTGTGTTTAATATTAAATTCTCTCCCATG
 GTTATCCCGTGAACAATCTTAAATGTCCTAAATTCTCATGACGTCATTAACACTCTATAACCAAACCTCTTGCCTG
 GGTTCTGTTTGTGTTAGTTGATGAAACAGAGTTCTAGAAGTICGTTCTGTTGAAAATTGAAAGCTTGGGTTG
 AGCTAAAGTTGGTTTATTACTGGGTCTGAGATTGAAGGATACTGGGTTGAGTTATATGGGAGGAGATGTTGGAGAGA
 TTGAATATGTTAATAGGATCAAGAAGATTATCTGTCATATCTGGAGAGATGTTGGAGAGCTAGAGATTGGGAGG
 CGCTCACCTAATAACAAAGCTTGTGACTAAATTGGTTAAGGATGGGAGCTAGAGATTGGGAGGTTGGCTGG

FIG. 27

GTTATGAAATGAGGCTAATTCACATTACCAATTGGCCATAATTGAACAGATAAGCAGAGGGTAATAATGTGGCTGAATG
 ACACAATTGGGGTTAATTCCTAAGAACATTCTGGCTATAGATCCCTTGGCTATTCAAGGACTCATTAGCTCAAGAAAGACCTTGCCCTGCCTATCTCT
 CGGGCTATGGGTTTGAAAACATGCTTATTCAAAGGACTCATTAGCTGAGCTGGATCTATGGAAACCACAGATACTTGTTCATATGGCTTATCTCT
 AATCTGAATAATTCTGGCTCAGAGCTGGATCTATGGAAACCACAGATACTTGTTCAGTTGATTTGCCATGGCTT
 ATTACATACGATACTCCACACACTTGTGGAGCCAGGCTGCAATTGTGCTGAGCTTGTGATGGGG
 ATTAAGTATGAAACTTGTCCATTGGGGAAAGCACCAGTGGAGACCAACTAGAAAATGTGGAGAGGGCATT
 AAGCTTCTGGATCAATAACAGGAAAAAAATCCACTCTATATCGAAACTAATACACTCTTACCTCTGGAGATGATT
 TAGGTACATTAGTATCGATGAAGCCGAGGGCTCAGTTCCTGAACCTACCAAGATGTGTTGATCACATCAACTCTAA
 TCCTAGTCTAAACCGAGGAAGCAAAGTTGGTACCTGGAGGATTATTCAGAACAGTCCGGAGAACAGAACAGA
 GTGAATTATTCTCGTCCCTGGTGAAGGTTCTGGTCAGGGTTCTGGTCAAGGTTATTATGTTICAAGAACCTTCAAAAGCTGTTGATCGTGT
 CATATCGAGATAAGCAACAAAGACTATTGGAGTGGTCAAGATCATGATGTCAATTCTGTAGGTATTGCCATGG
 GCTCGAGCATAACCCCTCGTGGAGCTGAGATCATGATGTCAATTCTGTAGGTATTGCCATGG
 AAATTCCAACAAGTTTACGTATAAGTGTACTGCTGCAAGAACAAATTCTGGCTCTTTCAGCACCATTGATGGGG
 TAACTGGAACTGCTAAGGGATTATGTGGTACAAGGTTACGGCACCCGGATGCATACTTCATTCAGAT
 CTTATGTCTAAAGCAATCGAAGGTCTCTTGTGGATTCGGCCACCGAGAACAAAATCTGATCAATTCCCCATCATTT
 TTGAGGGCAGAGCAAATGAGATCAAAAGTGTAGTGTGCTGCAAGGCTCAAGGGCAATTGCTGCCAGITCACAAGGCCAATTGCTGCCAGGGAAAGGAATT
 CGCACACAGTTTAACTCTCAATCCATCGAAACAGACGAGAGGGACTTGTGTCCTAGCCAATTCTCTGAAGTGCAGCATGACGATAACAAA
 CTATTCACCGGCAAGACATCGCCCTTAATCGGAAAGGCTCCATCCCCAGCTCTGGCTGAGAACATATTCAATTGCTA
 ATGGGAATGTCGAGTGTGAGAAAGCTACTCCGTCTAAACTCAAAACTCGGACAAACGTTACTGAGATCCGAAATGAACACATCAGACTCTTGTTGAT
 TCCTCCATATTCTCTGCTCCAAACTCGGACAAACGTTACTGAGATCCGAAATGAACACATCAGACTCTTGTTGAT
 GTGAAGAACGGATCACTGCGGAAGATAAGTCCCATAGAAACGGATCAAGAGACTGTTGGGGAGAACAGATAAGGTATGT
 ACTCTAGTCCAGAGAGTTGGAGGCTTACCTGTTCAAAACCAAGATGGTGAAGGCTAGCCAAATTGTTCAACCTGATGGACCA
 TGTAGTCACCTCTGAGGGTCTGCTGGGTTCAAGAACGTTCTCTTACCCCTAAACCAAAATGGGAGAAATCACCCCTC
 TCTCAGAAAACACTCGTCTTACACTGGAGGTAATACTGCTCAGGATCAAGTGGCTGAGATAAGAATATCATGTTGAGC
 TTCTGGTAATGAAATTGATGACCCGGAAATTGATGTCCTGGTACAAGACTGATGTTGAGAACAGGGCTTCT
 TTCAGATCTCAATGGTTCCAATGAGCAGGAGAAACTTATGATAAAGATCCCTCTCAAGGAAACTACTACCCA
 ATGCCCATCTCGCATTATCCAAGGATCCAAATGGTCAGAGATTCTCGTCAATCTCGTGCACTCTCGGTGTTG

FIG. 27 Cont.

CAAGGCCTCAAAGAGGGTTGGGAGATTATGGTGGACAGACGGTTGGTGTGATGACGGACGGGTCTAGGGCA
 AGGTGTGATGGATAACCCGGCGCAATGACCCTGGTATTTCACCTTCCTGGGAATCTAACATTCTCAAGCAGACCCCT
 GCTTCCAACACTAACCCGAGGAACCCCTCGCTCTCACCTCATAGGTGCTCACTAAACTACCCATAAAACA
 CATTCAATTGCCAAGAACATATCTGTGCGTGTTCACCAATACGGTTCCTTAGCCAACTAGCCAAACC
 GTTACCCATGTGACCTCCACATGTAAATTCAAGGGTCCCTCGTCCATCCAATTACTCTCAGCAATTGGAAAGAAC
 AAGCCAAGGGTCGCTTATCTCAATTAGCAGAGCTGGATTGCCATTAGCAGCTAAAGAACAGTAAGTAAACT
 GCACAAAGCAAGGCTAAACTTCCGACATGGTCAAGATCTTGAGCTTGGAGATTCTTGGTACGGATGACCAAGGACT
 TTCACTGAATCTCTGGAGATATGGAGATTCTCCATGGAAATACGAGCTTATAAGGCTTACAAGGTGAACTGGACCT
 CCACCGGAAGGACGTGTCGGATCTCACCAAGTCTCTCTTACAATCTCTTACAAGTCTCTTACAATCTCTTACA
 CCTGGTGAAGATCCGGTAGACTCCGCAAAAATCACCAAGTCTCTTACAATCTCTTACAATCTCTTACAATCTCT
 GAATAATGTGTGAGTAGTCCAGATAAGGGATTAGGGTCTATAGGGTCTATGTGTGAGCATATAAG
 AACCCCTTAACTCTTAAATTCTAACTCAATAAAATTCTAACTCTTAAATTCTAACTCAATAAAATTCTAACTCAATA
 AGAGACCTCTTAATTAA

FIG. 27 Cont.

AGATCTAATCCAATTACGATAACGCCCTGGGTACACTTGATTTCAGTGGTTACATATATCCTGGTT
ATATGGCTATCTTAAGGATCTGCACAAAGATTATTTGGTGTGTTGATGGGCTTGATGGGATTTGATATGATA
CACTCTAATCTTATGGAGATAACGCCAGGATTATACTCAGTAAGACAATTCAAATTTCAGTGGTCAAAACTCGTTA
TCTTTCATTCAGGATGAGCCAGAATCTTATAAGAATGATTGCAATCGAGAAATACTGTCGGCGATA TGCCCTT
GTTGGCTCAATATCTACATATCACACAAAGAACATCGACCGTATTGTACCCCTTTCATAAAAGGAAAACACAATAT
GCAGATGCTTCTCCACATGAGTAACATATGGTATTCAAAAATGGCTAAAGAAAGTTGATAACAAATTGAC
AACTATTCACATTTCTGTTATATAAAATTCAAC
AACTCTTATTTGGTATTGGCTAAGGCCAGCTACAGTGGGTACCGTGGGGTACCATATAGGAAGGTAACAAAAA
TACTGCAAGATAAGCCCTATAACGTTACCGCCTCTCCTTACCGAAGAGATAAGATAAGACCCACCCCTGCCACG
TGTCACTCGTCATGGGGTTAATGATAAGGGATTACACCCCTCTATGTTTGTGGACATGTCATGTAATGTCAT
GAGCCACAGGATCCAAATGGCCACAGGAACGTAAGAATGATAGATAAGATTTGATTTGTCATGCAAGCTAG
ATTATAAAGGTGTGATCAAAATTCAACTTGGATAGATAAGTCAAAATTCAACTCATTCCCTAAGTCCATTTCCT
AAACCATTGG

FIG. 28

FIG. 29

TGGTACAAGATTACGGCACCCGGATGCAACTTCATTGCAAGAACCTTCAGATCTTATGTCTAAAGCAATCGAAGT
 TCTTCTTGGGATTCGGCCACGGAGAAAGAAAATCTGATCAATCCCCATCATTTCAGGGAGGAAATTGCAACAGTTAATCTTCAATC
 AAGTATGATGCTGGCCAGTTACAAGCCAATTGCTGCCGGGAAAGGAAATTCTGATCAATCCCCATCATTTCAGGGAGGCAAAATGAGATCA
 CATCAGAACAGACGGAGGGAGGGTGGTGACGGTTGGTAAACCGCGCTGAAATCTCGGTTGGACTCAAACCTG
 GACTTGTGTCCTAGCCAATTCTCTGAAGTGCAGCATGACGATAACCAAACTATTCACTACCGGAGACATCGCCCT
 TACTGGAAAGGCTTCCATTCCCAGCTCTGGTCTGAGCTTGTGAGTTGACCCATTCCCTGCTCCATATTCCCTGCTCCAAACT
 CTACTCCGCTTAACCTCAAATACGCTTCTCTGAGATCGAGATCGAACATCAGACTCTGTGTTGATGTGAAGAACGGATCAGTGGGAAG
 GGACAAACGACGGTTACTGAGATCGAACATCAGACTCTGTGTTGATGTGAAGAACGGATCAGAGAGTGGGAAGCT
 ATAGTCCATAGAAACGGGATTCAGAGACTGTTGTGGAGAAAGAGATAAGGTATGTTAGCTAGTCCAGAGAGTGGGAAGCT
 ACCTGGTCAAAACCGATGGTGAAGGCTCAGCCAATTGTCACCTGATGGACATGTTAGTCACCCCTCTCAGAAAACCTAGTCACT
 GGTCAAGAACGTCCTACCCCTAAACCAAATGGGAGAAATCACCCCTCTCAGAAAACCTAGTCACT
 GGAGGTAATACGGCTTCAGGATCAAGTGGTCAGAGATAGAAATCATGGTGAAGCTTCTGGTAATGTGATTGTGACC
 GGGAAATTGGATTGTCGGTACAAGGACTGATGGTGACAACAAGGCTCTCTATTCAAGATCTCAATGGTTTCCAAT
 GAGCAGGAGGAAACTTATGATAAGATCCCTCTCAAGGAAAATCTACCCAAATGCGCATTTATCCAA
 GGATCCAATGGTCAGAGATTCTCCGTGCACCTCGTCAATCTCGGTGTTGCAAGGCTCAAAGAGGGTTGGTTGG
 AGATTATGGCTGGACAGACGGGTTGGTCTAGGGCAAGGGTGTGATGGATAACCGCGCAAT
 GACCGTGGTATTTCACCTCTCTGGGAATCTAACATTCTCAAGCAGACCCGTGCTTCAACACTAACCCGAGGAAC
 CCTTCGCTTCTCACCTCATGGTGTCTCACTAAACTACCCCATAAACACATTGCCAAGAACGGCAAG
 ACATATCTGTGGTGTCTCCACAATAACGGGTTCCATTAGCCTAACCGTTACCATGTGACCTCCACATTGT
 AAATTCAAGGGTCCCTCCAAATTACTCTCAGCAATTGGAAAGACAAGGCAAGGTTGGCTTATCCCT
 AATAGACGAGGCTTGGGATTCACTTATGCCATAAAGGAAGACAAGTAAACTGCACAAAGCATGGCTATGAACCCAG
 TAAACTTTCGACATGTTCAAAAGATCTGCAAGCTTCAAAAGGTAACCAACTTCAGTGAACCTGCTGAAGAAGTC
 TCTCCCATGGAAATAACGGAGCTTACCTCGAGATAGTTCACAGCCACGGGAAGGACGTTCTCGACCTGCACCTGCTGAAGAT

FIG. 29 Cont.

GGCGCGCCTCGGGGATCGCAGATCTCATTACCGTTAGAAGCCATAGTAAACACTAAAGCTTGTGTCGTTAACATT
 TAGTCATTTCACATTGTGGTTCTACATTAAATGAATTTCATAATGCCAATACAGAATTAAATCAAAATTGTT
 TGAATTATGCTAAACATGTAACATACGTAATTCTCGCCTTGTGTGTATTAAACTTGAAAGTTATCATAAAGAACCC
 ACAAAATACACTAGTAATCTATGAAAGGCAGGTGGCAACACAACAAAGAGTATCTAAGATTTCATTTGTGACTA
 TAGGAATATAATCTCTTATCTGATTAAATGAATTCCACATGTTCACCTCTCATTTGCCACAAAGATCACAAACTTT
 ATCTTCACATATTCACAACTTGTATATCCACCAAAATTCAATTCTCTTCACTTAGCCCCACAAAATACTTTGTCCC
 CTTATTGCCACCTTGTATTAAATTCTTGTGGAGGCTAAGTGTTCATATTCTCTTCAAAAAACA
 AAAACAAAAAGAGAAAGAAAACCATTACTGAAGGGCCAAAGCGGTCTGCTCTGAGCTGCAGAAAGAAAGTGA
 CTGCAACCCCTTCCATTACTGAAGGGGAGGAGGAGGAGGACCCAGGTGAAATTCTCGTCAAGAGCACC
 GTTCTGGGACCCGTCAAAACCTCTCGTCAAGAGCACCAGGTGAAATTCTCGTCAAGAGCACC
 TGAAAATTTGGGGATGATTAAATAACAAAGGGTGTACCTCTGGGAAACACTGACTTGGGGACCATGGCTCT
 AAGGTGCAATAGAAGGAACCATTTTCCGCTAGAAATACGGGATCTGTTCCCGGATTGGCAAAAGATCGTGTGGTT
 ATCGTCTTGTGATAATGGCTCAAGTATTTCGACTGAAAGGGTAAAGGTATAAA
 GTGAGACATTGGTGTGATGGTCAATGATGGTTACCTTGAAGAGATGAATAGGATTGGGAGGAGTATTAAAGTTTG
 TCAAGTGAAACAGATTTCGCCCCATTCGGCTCATATATATCGTACTAGCTTCCGGGTGTGACCTGTAATGGGA
 TGTAAGAACAAAGGGTGTGATGAGGCCAAAGGGCAATTGTGAAGGTAATCCGTGATCAGTATGGGAATCATGGGTCTCCGA
 AGATTGGTATCTTGAAGGCATCACTGGTGTGGATGATGAACACTGTATGGGATGGGTGGATGATGA
 TGAGGGCATATCCTTTCATGGAAAGATCAATTTCGTTCTGCTTAATGCCATACAGACTCTTACG
 AGGCTGAAACCCGGCAAAGTGTCTGACTGTTGCTAATTAGCACCGTCTGATGTGAAGGTCAAGAGGAGAAG
 GGCTTGAAGTGTGGTGTGAGAGAAATGGGAATTGTGTGGTATTCTTTAATAGAAGTGTGTGGAGAATATTCA
 TCAGAAGGCCAAAGAGAGGTTTGTGATTACAACCTGGGATAATAACGATGTGGGCAACGGGTTTCCCGTGC
 TTGGGTTCCCCGGTGTACACATGGCAGGGCCTAGGACTAGTGCAGGGTACACTTTGGAAAATGTGGGTTCATCAAG

FIG. 30

GTAGAGGAGATGAGGGTGAATTGCATAATGGGGTCGATAACATAGAAGTTAAGGAAACAGATAAACAGATAAAGTTGTGAA
CTAAAAGAAGGGATGGGGAGTTCGGGGTATAAGCATCAAGCGGGTTATAAGCCGGTTACAGCAGTGGCATCTC
TGGGGCGATGATAAGGGAACCGACATTAACTTAACTGTTGGATTGCAACTATGTTACAGCAGTAGCAGTGCATCTC
CATGAAACGGATCCGCTAGAGTCCGCAAAAATCACCAGTCTCTACAAATCTATCTCTCTATTITCTCCA
GAATAATGTTGAGTAGTTCCCAGATAAGGGAAATTAGGGTCTTATAGGGTTTCGCTCATGTTGAGCAATAAG
AAACCCCTAGTATGTAAATACITCTATAATAAAATTCTAAATCTAAACCCAAATCCCCGG
AGAGACCTCTTAATTAA

FIG. 30 Cont.

AGATCTCATTATAACCGTTAGAAGCATAGTTAAAGCTTGTGTCGTTAACATTCTAGTCATTTCATTGCTTAATTCTAGTCATTGCTTACATTGTTGGG
TTCTACATTATAATGAAATTCTAAATGCAAAATCAGAATTAAATGCTAAATTGCTAAACATGTAA
CATACGGTATATCTCCGCCTTGTTGTTGTTGATTAAACTTGAAAGTTATCATAAGAACCAAATACACTAGTAAATCTAA
TGAGAAGGCAGGTGGCAACACAACAAAGAGTAICTAAGGATTTCATTGIGACTATAGGAATAATAATTCTCTTAT
CTGATTAAATGAAATCCACATGTTCACTTCATTGTCCACAAAGATCACAACCTTATCTCAAAATTCAACAACTTG
TTATATCCACCACAAATTCAATTCTTCACTAGCCCCACAAAATACTTGCCCCCTTATTTGCCACCTTTGTAT
TTAATTATTCTTGTGGAGCTAACATTCTCTCAAGTGTICATATTCTCTTAAACAAAAACAAAAACAAAAAGAGAAGA
AAACCATGG

FIG. 31

CCATGGCAGAGGGAGGAGATCAGTGGTAGGCAGCAGCAAAATGGAGGTACTGCAACCCCTCCTATTACTTGAA
 GCGCCCAAAGCGTCTTGTCTCTGCTCTTCATCGTTTCGTTGTTCTGGGACCGTCAAAACTCTC
 GTCAGAGGCCAGGGTGAAATTCCTGAGCTGCAGAAAAGAAGTGACTGATTGAAAAATTGGGTGATTGATTIAA
 ATAACAAACAGGGTACCTCTGGAAAACCTGACTGGGACCATGGCTTAAGGTTGCATAAGGAAACCATT
 TTCGCTAGAAATACGGATCTGTTCCCGATTGGCAAAGATCGTGTGGTTATCGTCTTGTATGTGCATAATCGG
 GCTCAGTATTTCGAGTCACAGTGGAAAGGTTAAGGTAAAGTGAATAGGATTGTGGAGAGTTAAAGTGTGATTTGTTAGC
 ATGATGGTTACCTTGAAGAGATGAATAGGATTAAAGTGTGACCTGAAATGTAAGGAACAGATTCTCGCC
 TTATTGCCTCATATAATCGTACTAGCTTCCCGGGTGTGACCCCTGAATGATTGTAAGGAACAAAGGGTGTGAGGGCA
 AAGGGGCATTGTGAAGGTAATCCTGATCAGTATGGGAATCATCGTCTCCGAAGATTGTATCTTGAAGGCATCACT
 GGTGGTGGATGATGAACACTGTATGGGATGGGTGGAAAGAGACTAAAGGACATGAGGGCATATCCTTTTCATIGA
 AGAAAGATCATTTCTGTTCTTAATGCCATACGACTCTAACATACAGACTTACGAGGCTGAACCCGCAAAGTGTCT
 GACTGTTTTGCTGCTAATTAGCACCGTCTGATGTGAAGTCAAGAGGAGAAAGGGCTTGAAAGGTTGGCAGAGA
 GAATGGGAAATGGGGTATTCTTTAATAGAACGGTGTGGGAGAATATTCAATCAGAAGGCAAGAGAGTTGGTT
 CTTTGATGATTACAACRGGGATAACGATGTGGCAACGGTTTCCCCGGTGTGACACATTTG
 CGAGGGCCTAGGGACTAGTGTGGGTACACTTTGGAAAATGTGGGTTGCATCAAGGTAGGGAGATGAGGGAGTTGCA
 TCGATAATGGGGTGTAAACATAGAACGTTAAGGAACAGATAAAGTGTGAAACATAAAAGAAGGATGGGGAGTTG
 GGTGATAAGGCATCAAGGGGTATAAGGGTTCTGAAGGTTGGGAGTTGGCGATAGGGACCGACAT
 TTATGGTGGATTGGCACTATGTATCGTTACAGCAGTGCATCTCCATGAAACGGATCC

FIG. 32

GGATCCGCTAGAGTCCGCAAAATCACCAAGTCTCTCTACAAATCTATCTCTCTCTCTCTCTCTCTCTCTCT
 TGTGAGTAGTTCCAGATAAGGAAATTAGGGTTCTATAGGGTTCGCTCATGTGTTGAGCATATAAGAAACCCCT
 AGTATGTATTTGTTGTTGTAAGGTTAAACTCTCTATCAATAAAATTTCTTAATCCTAAACAAAATCCCGAGAGACCT
 CTTAATTAA

FIG. 33

Lanes: 25μg total protein each
1, MGR48-p antibody control
2, TmXyl-GalT
3, GalT
4, SSNN control

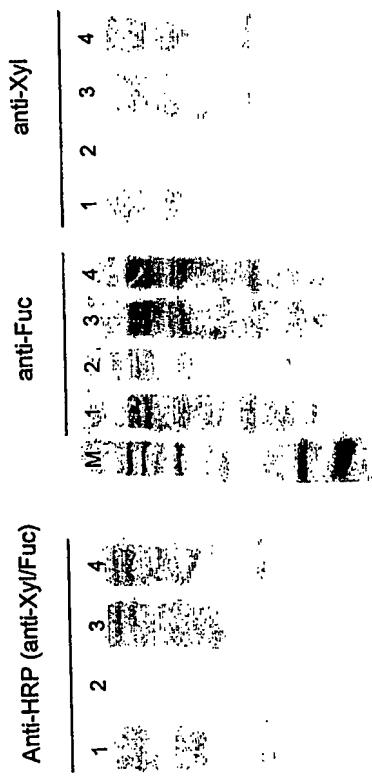


FIG. 34

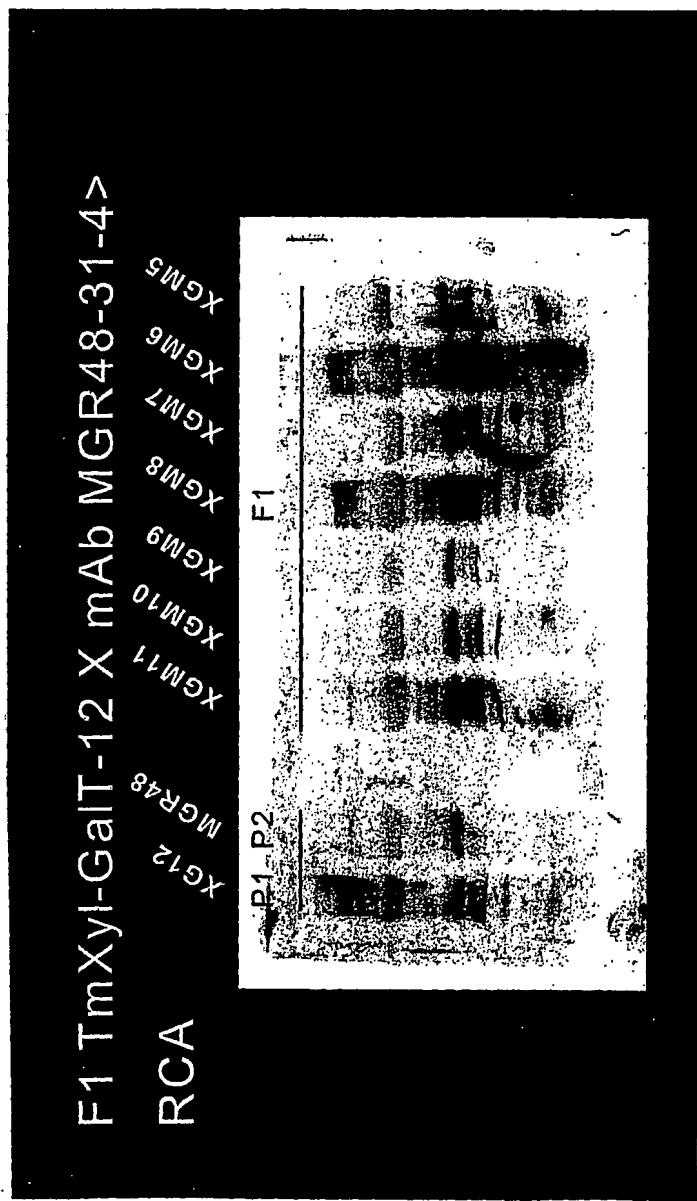


FIG. 35

- Lanes 1, 2 & 3: 300 ng MGR48 IgG
- 1, MGR48 hybridoma
 - 2, MGR48 tobacco
 - 3, MGR48 TmXyl-GalT

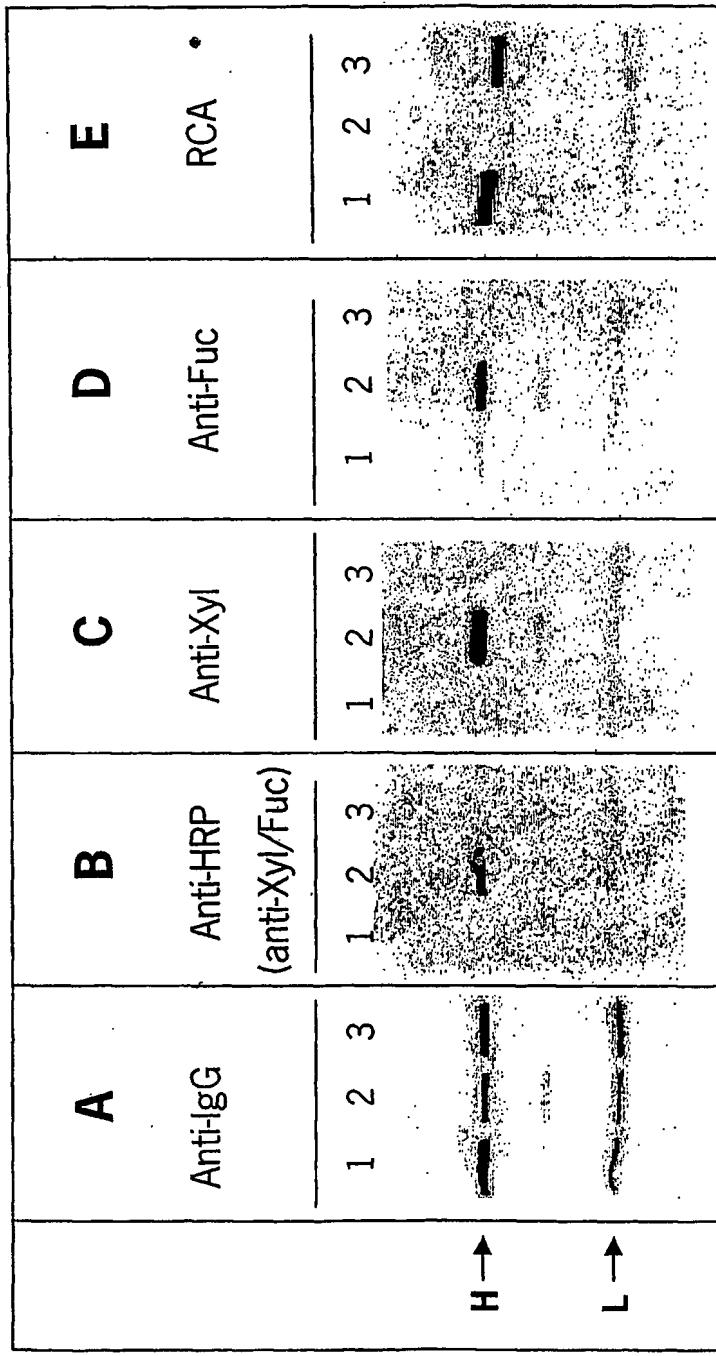


FIG. 36

ATGGGGCATCAAGATGGAGACACATTCTCAGGTTCTTGTATACTGTTGCTGTTGCTGGTGTGACATGAAGCATTCAAACTTCCCTCACT
CACACCGTCAAGAGCCGAGACGAGCCAACTCCGGATCAATGCCCTGCAATTGGACAGGGACATCGACACCGTGGGATATAACCAACT
TTTGATTTCAAGGACAAGTCCCTCAACTCACCAAGGAAATTGGTGGTTGCTGAGGATCGGTATGAAAGAATTCAAACTACAGCACTACGGGCT
AGACTGAAGGTAATCGTTGCTTCAGTCAACTCACCAAGGAGGGATGGCTGAAGAGCATTGAAACAGTACTTCAGGTTGCTGACCT
AACAAACATAAGTGAACAAACTTATCAAAGAAGGTGCTCAGATCACGACGGGGGGTGGGTGATGCCGGACGAAGCCCTGCACGCATATC
GTCAAAACAAAAGGCATTGAAAGAAACTTATCAAAGAAGGAAACTAACTCGGGTAAAGACGGATGCTTATGCCGAAGAACAGGATGGCT
TATGCGCTAATTGACCAGTTATTGAAAGGACATCACTGGGTAAAGAAACTAACTCGGGCTCATCCGGTCACTGGGCTGGGGCTGGGGAG
CACGGGCCACTGGCTTACCTGCTGAGCCATTATCAAAGAAGAACATTATGCGTGGAAACAGTGGCTGGGGAG
CGACAGATTGAGGAGTTTACTGGCTGGCGAGTTGGCTACTACGAAAGCCGTTGGCTTACGAGGGATTACGACCGTATGGGTTCCCTGACT
AGCACGTTGGCCCGCACCTCAATTGTCAGTTGACTCAAGGAAGATTCCGGAAATATTCTGAATAACAGCTAAGCACGAAGACATC
ACGGAAACACAACCTGCACAGCAAGGCAAAGACTTTGATAAGGGATTACGACCGTATGGGTTCCCTGACTCCACACAGTGGCTGGGCT
GGAGAGCAGACTCAGATAACGAGTACAGCGTCAGGTTGATGCCCAATTACGTCAAATTACGTCATAATGAAAGTTAACTACATCAATGCT
TTCAAACGCTGACGTACAGTTCGGAAACTCCTCTCGATTACTTTAAGCCATGAAAGAACATCAAATAACCCAGCTTAAGGGAGATTCTTC
GTTTACTCCGGATAATTTCAGCGGAAGGTTAACACGGTACTGGTCAGGTTACACTAGACCCCTACCAAAATACTCCGGCTCAGTTGCAA
CACCAACTGGCGATGGCAAGAGATTATTCAACCGTCAACTACGAGATGGGTCTGTTCAACACGTTGCAACTACGGATGGGATTTCT
TTAGAAAAAAATCTTACGAGCAAGCTTATCTATGCTCGACGGAAACTTGGGTCTGTTACTGCAAGGTTCAACAGTCTGTATCACT
ATGCAAGGATTACGGAAACCAACTGTTCAACAGTCTGTATCACTGCAATCCGGCTGCAAGGGGGCTCACCCATCATGTTGCCCTGACCA
TTGCACTCGCAGAGCATTACAAAGCGAGGTGAGTGGGAAACTTACGGAAAACCGGCCAAGAAGCTGCAGTGTCTTCAATTGACAAGGAAA
GTTATACTTTAATCCGGTGGCTGAGACTCGAAAGTGGTCACTGTTAGATCCAACACGTTCAACATCCGGTGTAGATAACACGGG
AAGCACGCTCTTCACTGAGATAATGCCAGCATACAATCCAAAGACAACGGCAAGAGTATCGTAAGCGACACCAGTTGACATAATGATA
ACCATCCCCCTCACCTCCATCTGTCAGAGGGACACCAACACTTCCACCAACTTCTGCAACAAACTCGCGTCATTCTGCAACAAACT
CAGAAATCCAATGTTGCTGGTGAATAAAATTAGAAAATGATGCCTGGTACATAAAACTCTGCACTGCTAAACTTCTGTTAATAGGAACACGGG
TTTCTGAGACAAGGCTATAGAAAGGACATCCGGAAAGAGAAACTGTCGTTGACGTACAATTGGCGCATATCAAAAGTGCCTAAAGACATTCTGGTGT
TACCTCTCATGCTCATACGACTCACCTGAAAGAATGTTCTGCAATCCCTACACTATCGAAACAACATGCAAGATGATAACATAATAGTG
TCCGGACCTATTCTACGGAAATCACGACCATGTTAGATTGGTCACACTTGGGATGACTGAGTTATTAGGATAATACAACTGGGACTCTGCGGT
GCTATTCTATTAGAGACCGATGTTAGATTGGTCACACTTGGGATGACTGAGTTATTAGGATAATACAGAGAGACTGAGTTATTAGGATA
GACATTCCCAGAATTACACCGATCAAGGGATTCCAGTACCTGCAAAAGGGTCAAACTGCAACCGCTCAAGGGCTCAAGGGCTCAAGGG
ACTACCATGGCGTGCCTGCAAGCAGGGAGACCCGGCTCACTCTGCTGACATAGGAACTAGGAAATAACTAGGAAATAAGGTTAGAA

FIG. 37

GTCATGCTCGATCGTCAACTCTTTATGATGGACTTCAGAGGAATCGGTGAAGGGAGTAGTCGATAACAAACCGACGACTTTCAGAAACTGGATTAA
ATTGAATCCATGCCAGGGCGTGACGGCGTAAGAGAGACACTAGTGAAGAGAGAAATTGTTAATGAACTGGTCTGGCCGGCCAGAAG
GAAAGCCCTTACCAAGTACCGTGCAGACTGGGACTACCTGAGCAGGTGTTCAATTACCCGGTGAACGTTACCTGGTGGACACTAGCGAGGT
GGCGAGATCGAGGTGAAGCCGTACCAAGTCTGCAGAGCTCCTGCAGAGCTCCCGCCCATCACCTGGTACCCCTGGCACCACATCACCGGACGACTGCT
GAACTCTCCCCAGCAACGAAAGCTACATGGTACTGCACCGACAGGATAACAGTGTGCGCTGTGGAGAGAACGCTCCAAAGTCTGGCCAAGT
TCGTCCAAAACCAAGGGTCAATGGTCTGAACATTCAAGAACATCACGGCAGTCAGCCTGACCGGCTGAAGTCACTCCGACCTCTCACAGGTCTGAGT
GACATCCACCTGAAACGCTATGGAGGTAAAACCTACAAGATCAGGTTAAGGACGAGCTTAA

FIG. 37 Cont.

MGIKMETHSQVEVYMLLWLSGVDMKHFKSSLTHTVKSREPTPDQCPALKESEADIDTVAIYPTFDFQPSWLRTRKEFWDKSFEDRYERIHNDTTRP
RLKVVVPHSHNDPGWLKTTEEQYFEWKTKNLIINNIVNKLHQYPNMTFIWTEISFLNAWERSHPVKQKALKKLIKEGRLEITGGWVMPEACTHI
YALLDQEIEGHHWVKTNLGVIPKTGWSIDPFHGATVEYLLDQSGLETIIQRIHYAWKQWLAERQIEEFYWLASWATTKPMSMIYHNQPFDIYSIK
STCGPHPSICLSEFRKIPGEYESEYTAKHEDITEEHLHSKARTLIEEYDRIGSLTPHNVLVPLGDDERYEYSVEFDAQYVNYYMMENYINAHKEI
FNADVQFGTPLDYFNAMKERHQNIPSLKGDEFVYSDIESEGKPAYWSGYTTTRPYQKTLARQFEHQRLRSAILFTLVSNYIROMGRQGEFGASEKK
LEKSYEQLIYARNLGLFQHHDAITGTSSKSSMQDYGTKLFTSLYHCIRLQEAAALTIMLPDQSLHSQSIIQSEVEWETYGKPPKKLQVSFIDKKK
VILENPLAETRTEVVTVRNTSNIRVYDTHKRKHVLYQIMPSITIQDNGKSIVSDTTEDIMEVATIPPLTSISYKLOEHTNTSHHCVIFCNNEQY
QKSNVFQIKKMPGDIQLENNAVLKLVNRNTGFLRQYRKDIRKRTVVDVQFGAYQSAQRHSGAYLEMPHYDSPEKNVLHPTYTNQNNMQDDNIIIV
SGPISTEITMYLPPFLVHTIRIYNVPDPVLSRAILLETDVDEAPPKNRETELFMRLOTDIONGDIPEFYTDONGFQYQKRVKVNKLGIEANYYPI
TTMACLQDEETRLLTNTNHAQGAAAYPEGRLEVMDRRTLYDDERGIGEGVVDNKPTEQNWLIESMPGVTRAKRDTSEPGEKFVNERRFGPQK
ESPYQVPSQTAADYLSRMFNYPVNVYLVDTSEVGEIEVKPYQSELQSFPPGIHLVTLRTITDDVILEFPSNESYMLHRPGYSCAVGERPKAKSPRF
SSKTRFNGLNIONITAVSLTGLKSLRPLTGLSDIHLNAMEVKTYKIRFKDEL

ATGGGCATCAAGATGGAGACACATTCTCAGGTCTTGTATAACATGTTGCTGGTTGCTGGCTCCGGAGCTCCGGGCC
GGAGGGGGCGCCGGCCGGCCCTCTAGGGCCTCCCTCAAGCCCCGGGACTCCAGGCCAGTCGTGGATTCTGGCCCTGGCCCCGGCT
AGCAACTTGACCTGGTCCCAAGTCCCCACACCCGGCACTGTCGTGCCCCCTGAGGAATCCCCGGCTATGGGGACTGGCTCTCTCC
GAGTTAACATGGCTGTGGACCTGGCATCATCATTCCGAAACGGCAGGACACTATTAATCAGCTGGCTAAGCTCCTCAATGTTGGCTTC
ACAAGGGGGCCATCATCATTCCGAAACGGCAGGACACTATTAATCAGCTGGCTAAGCTCCTCAATGTTGGCTTCAGAACGGCTTGAAAGGACTAT
GACTATGGCATCTATGTTATCAACCAGGGGGAGACACTATTAATCAGCTGGCTAAGCTCCTCAATGTTGGCTTCAGGCCACGGCACATTCCGGT
GACTACACCTGGCTTGTGTTAGTGAACCTGGACCTCATCCAAATGAATGACCAATAATGCGTACAGGTGTTTCAGGCCACGGCACATTCCGGT
GCAATGGATAAGTTGGATTCACTTAGTGTCTGGGGAGAAGATGACATTAAACAGATTAGGGCATGTTAGGCAATGTTAGGGCAATTCAGGGT
CCTAATAATTATGGGCTGGGGAGAAGATGACATTAAACAGATTAGGGCATGTTAGGCAATGTTAGGGCAATTCAGGGT
GGGAGGGTGTGGCATGATCCGCCACTCAAGAGACAAGAAAATGAACACACAAGGAGACAATGGCAATTGACACACAAAGGAGACAATGGCT
TCTGATGTTGAACCTAACCTAACCCATTGTTACAGGAGATAACCCAAATCACAGTGGATGTTACAGGAGATAACCCAAATCACAGTGGACATCGGGACACCCGGCAAG
GACGAGCTTAG

FIG. 39

NGIKMETHSQVFVYMLIWLSGVDMQSSGELRTGGARPPPLGASSQPRPGDSSPVDSPGPASNLTSVEVPHTTALS LPACEESPLLVGPMLI
EEENMPVDELLVAKONPNVKMGGRYAPRDCVSPIHKVAIIIPFRNROEHLKYWILYLPVLRQQLDYGIYVINQAGDTIENRAKILLNVGEQEALKDY
DYTCCFVFSVDLIPMNDHNAAYRCFSQPRHISVAMDKEFGESLSPVQYEGGVSAALKQQFLTINGPNNYWGWGEDDDIENRLVERGMSISRPNNAV
GRCRMIRHSRDKKNEPNPQRFDRIAHTKETMISDGGLNSLTYQVLDVQRPLYTOITVDIGTPSKDEL

FIG. 40

ATGGGCATCAAGATGGAGACACATTCTCAGGTCTTGTATACTATGTTGCTGGTGTGGCTGGACATGGGACACAGATGCCCTGGCTGGCTGTA
 GTGGTATGGCCTGCAAGTCAGTCAGTGGATCTGATCAAGCTGTCAAGAGCAAGTCATTGAGCTATAATCAAACTCCCGTTGCTTAACATATCAGTCTTATTT
 ATATCTCAGGATGGATCTGATCAAGCTGTCAAGAGCAAGTCATTGAGCTATAATCAAACTCCCGTTGCTTAACATATCAGCAGCTGGATTGAACCAGTGGTC
 ACTGAAAGGCCTGGCGAAACTGACTGCGTACTACAGATGCACGTCACACTACAAGTGGCAGCTGGCAGCTGGCAGCTGGCTGAGCTAGTCTCATGGATAGGGATAAAACCAATTAGTCGA
 GTGATATACTAGAAGATGATAATGGAAATTGCTCCAGACTCTTGATTACTTGAGGCTGAGCTAGTCTCATGGATAGGGATAAAACCAATTAGTCGA
 GCTGCCTTCATCATGGAAATGATAATGGACAGAAGCAGTGGCTGAGCTGGCTGAGGATTTTCTGGCTTGGGCTGGGATG
 CTCAGAGAGATGGACTTGGAATGGATGAGTATCACCAAAGTGGCCAAGGGCTTACTGGGATGATTGGCTGAGACTAAAGGAAAACCAATAAGGGCCCAA
 TTCAATTGACCGGGAAAGTCAGTGTAGAACATAACATGGTGAACATGGCTAGTTGGGACAGTTTCTGAGCTAGTCTGGCTTAAAGCTATAAGGACTATAAGGCT
 AACGATGTGACGGTTGACTGGAAAGCAAGGACCTGGGATAACCTGACAGAGGGAAACACTACCAAGTACTTCTGGCTTAGTGAACAGAACGACAGA
 CCAATTCAAGGTCTGACCTTGCTTAAGGATGATGGTCGATTCGGTATAAGGCTCAAACATAACGCTTAAAGGATGATGGTCGATTCGGTATAAGGCTAAAGGACTGTA
 GGGAAATTGGTATATTGAAGAATGGAAAGGATGGTGTGGCTCGAACAGGCTATAAGGCTAAAGGATGATGGTCGATTCGGTATAAGGCTAAAGGACTGTA
 TTCTGGTGGCCAGATCTGTAAATGGAGCTTGGAAATTGCAAGGCTTGA

FIG. 41

MGIKMETHSQFVYMLLWLSGVDMGQMPVAVVVMACSRADYLERTVKSVLTYQTpvASKYPLFIQDGSDQAVKSLSYNQLTYMQHLDFFEPVV
 TERPGELTAYKYLARHYKWAQDQLEFYKHKFSSRVIDEEDMEIAPDEFDYEEAAASLMRDKTIMAASSWNDDNGQKOFVHDPYALYRSDFEPGLGWM
 LKRSTWDDELSPKWPKAYWDDWLRILKENHKGRQFIRPEVCRTYNFGEHGSSLGQFFESQYLEPIKLNDFVDTWKAKDLGYLTEGNYTKYFSGLVRQAR
 PIQGSDIVLKAQNIKKDDVIRYKDQVEFERIAGEFFGIEEWKDGVPRTAYKGVVVFRIQTTTRRVFLVGPDSVMQLGIRNSKDEL

FIG. 42

54/57

ATGGGCATCAAGATGGGACACATTCTCAGGTCTTTGTATACTGTTGCTGTGGTTGCTGGTGTGACATGGCTCTAAGGTTGCATAGAAGGAAC
CATTTTCGCCCCGATTCGGAAATACGGATCTGGTTCGGTTATCGTGTGCTTCAGTATTTTCGA
GTCACAGTGGAAAGTTGTGCGAAGGTAAAGGATAAGTGGACATTGTGATTTGTTAGTCATGATGGTTACTTTGAAAGAGATGAAATAGGATTGTT
GAGAGTATAAGTTGTCAAGTGAACAGATTTCGCCCTTATTCGCCCTCATATAATCGTACTAGCTTCCGGGTGACCCCTGAATGATTTGT
AAGAACAAAGGGTGAATGAGGCAAAAGGGCATTGTGAAGGTAATCCTGATCAGTATGGGAATCATCGGTCTCCGAAGATGTATCTTGAAAGCATAC
TGGTGGGGATGATGAACACTGTATGGATGGGTGGAAAGAGACTAAGGACATGAGGGCATATCCTTTCATTGAAGAAAGATCATTTCCTGTT
CTTAATGCCTATGTAACATACAGACCTTAAGGCTTACAGGGCTGAAACCCGCAAAGTGTCCCTGACTGTTTGTGTGCTAATTAGCACCGTCTGATGTGAAG
TCAAGAGGAAGGGCTTGAAGTTGGTGCAGAGAGAATGGGAATAGTTGTTAAATAAGGTTCTTCCCGTCAACGATGGTGTGTTCCCGGTGTAACATTCAGAAG
GCAAGAGAGTTTGTCTTGTGATTACAACCTGGGATATAACGATGGTGTGCTCAAGGTAGAGGAGATGAGGGTGTATAAGGCTGATAATGGGTCTGTAACATA
GGCCTAGGACTAGTGCCTACACTTGGAAATGTGGGTACACTTGGTGCATCAAGGTAGAGGAGATGAGGGTGTATAAGGCTGATAACGATGGGAGTTGGGTGTAAGG
GAAGTTAAGGAAACAGATAAAAGTTGTGAACATTAAGGAGATGGGAGTTGGGTGATAAGGCTGATAAGGCTGATAACGATGGGAGTTGGGTGTAAGG
TGGGAGGGTGGGGCATAGGGAACATTATGTTGGATTTGCCACTATGTTGAAAGGCTGACATTATGTTGGATTTGCCACTATGTTGAAAGG

FIG. 43

MGIKMETHSQVFVYMLWLSGVDMALRLHRRNHFSPRNTDLFPDLAKDRVVIVLYVHNRAQYFRVTVESLKVKGISETLLIVSHDGYEEMNRI
ESIKEPCQVKQIESPYSPIYRTSEPGVTLNDCKNKGEAKGHCEGNPDQYGNHRSPKIVSLKHHWWMMNTIVWDGLETKGHEGHILFIEDHFLF
PNAYRNIIQTLTRIKAFCAPCPCDFAANILAPSVDKSRGEGLSVAERMGNGVYSENRSVMENIHQKAREFCFDDYNWDITMWATVEPSEGSPVYTLR
GPRTSAVHFGKCGLHQGRGDEGDCIDNGVVNIEVKETDKVUNIKEGWGVRYKQAGYKAGFEWGWWGDDRHLCLDEATMYRYSSSASPKDE

FIG. 44

FIG. 45

MLKKQSAGLVLWGAILFVAWNALLLFFWTRAPGRPPSVSALDGDPA SILTREVDMQSSGELRTGGARPPPPLGASSQPRPGDDSPVVDSPGP
SNLTSPVPHTTALSPLACEESPLLVGPMLIEFNMPVUDLVEAKONPNVKMGGRYAPRDCSVSPHKVAAIIIPFRNRQEHLKYWLYYLHPVQLRQQL
DGYIVVINOAGDTIENRAKLLNVGFQEALKDYDTCFVFSVDLIPMNDHNAYRCFSQPRHISVAMDKEFGESLPLPVQYEGGVSA LSKQQFLTINGE
PNNYWGWGGEIDDIFENRLVFRGMSLSRPNAVVGRCRMIRHSRDKKNEPNPQRDEFRIAHKTETMLSDGLNSITYQVLDVQRYPLYTQITVDTIGTPS

FIG. 46

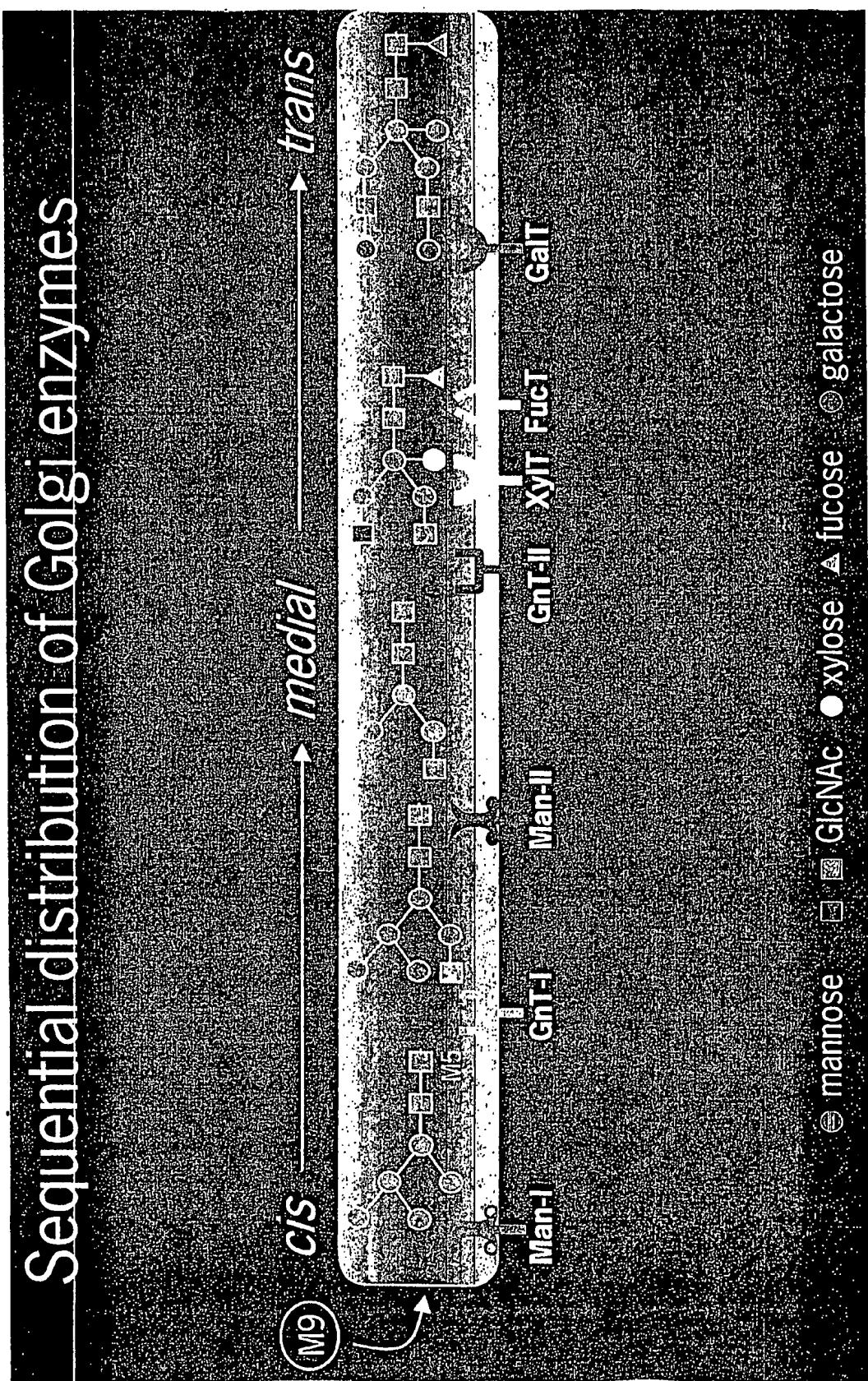


FIG. 47

Relocalization of Gait by CTS-region swapping

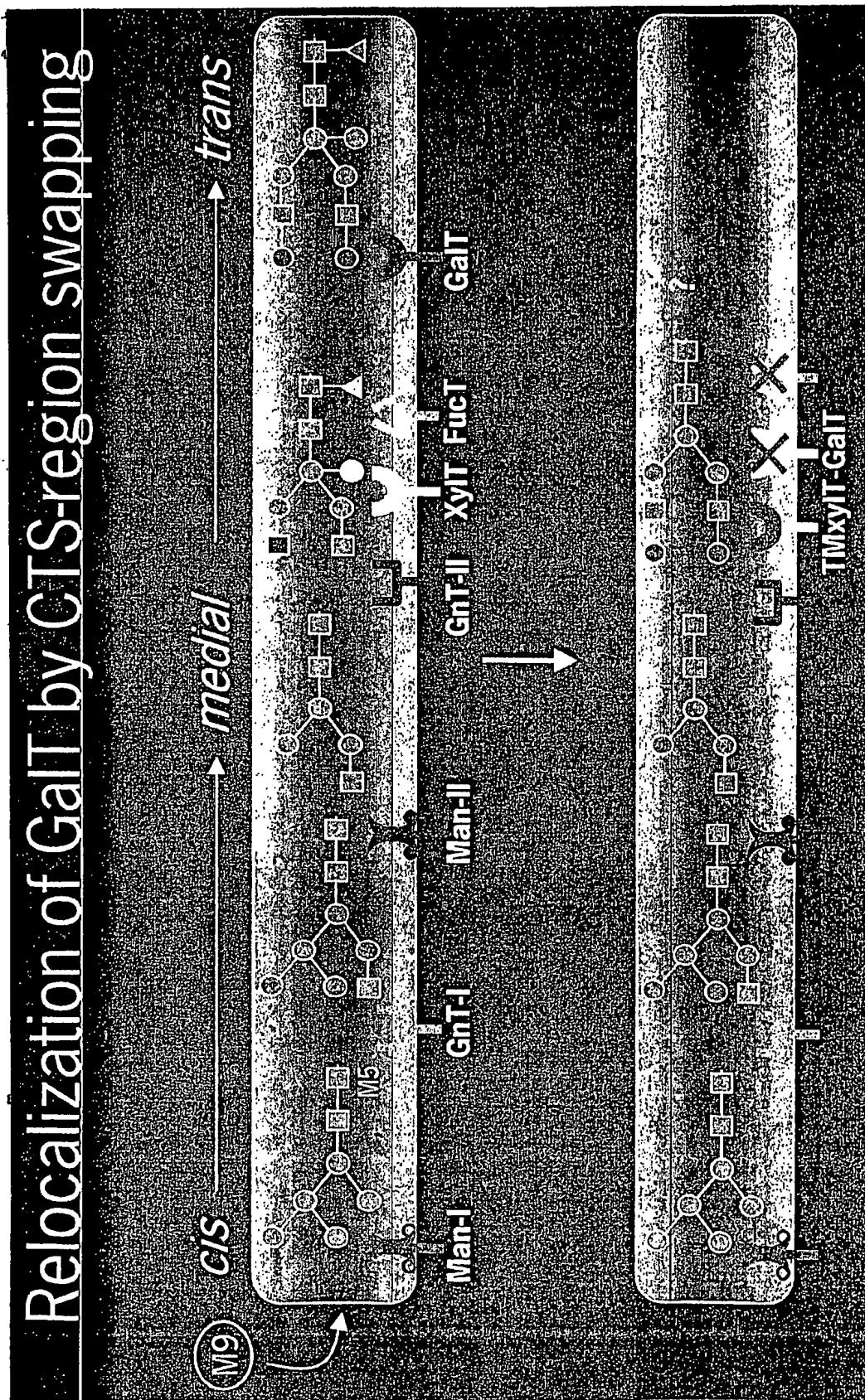


FIG. 48

SEQUENCE LISTING

<110> PLANT RESEARCH INTERNATIONAL BV
 BAKKER, Hendrikus A.C.
 FLORACK, Dionisius E.A.
 BOSCH, Hendrik J.
 ROWENDAL, Gerard J.A.

<120> Optimizing glycan processing in plants

<130> 62861A - P033486W0

<150> US-60/365,735
<151> 2002-03-19

<160> 59

<170> PatentIn version 3.2

<210> 1
<211> 1197
<212> DNA
<213> Homo sapiens

<400> 1
atgaggcttc gggagccgct cctgagcggc agcgccgcga tgccaggcgc gtccctacag 60
cgggcctgcc gcctgtcggt ggccgtctgc gctctgcacc ttggcgtcac cctcgttac 120
tacctggctg gccgcgaccc gagccgcctg ccccaactgg tcggagtctc cacaccgctg 180
cagggcggct cgaacagtgc cgccgcaccc gggcagtcct ccggggagct ccggaccgga 240
ggggccccggc cgccgcctcc tctaggcgcct tcctcccagc cgcgccccggg tggcactcc 300
agcccaagtgc tggattctgg ccctggccccc gctagaact tgacctcggt cccagtgc 360
cacaccaccc cactgtcgct gcccgcctgc cctgaggagt ccccgctgct tgtggggccc 420
atgctgattg agtttaacat gcctgtggac ctggagtcg tggcaaaagca gaacccaaat 480
gtgaagatgg gcggccgcta tgccccccagg gactgcgtct ctccctcacaa ggtggccatc 540
atcattccat tccgcAACCG gcaggagcac ctcaagtaact ggttatatta ttgcaccca 600
gtcctgcagc gccagcagct ggactatggc atctatgtta tcaaccaggc gggagacact 660
atattcaatc gtgctaagct cctcaatgtt ggcttcaag aagccttgaa ggactatgac 720
tacacctgct ttgtgttag tgacgtggac ctcatccaa tgaatgacca taatgcgtac 780
aggtgtttt cacagccacg gcacatttcc gttcaatgg ataagtttg attcagcccta 840
ccttatgttc agtattttgg agggtgtctct gctctaagta aacaacagtt tctaaaccatc 900
aatggatttc ctaataata ttggggctgg ggaggagaag atgatgacat ttttaacaga 960
ttagttttt gaggcatgtc tatatctcgcc ccaaattctg tggtcgggag gtgtcgcatg 1020
atccggccact caagagacaa gaaaaatgaa cccaaatcctc agaggtttga ccgaattgca 1080
cacacaaagg agacaatgct ctctgatggt ttgaactcac tcacctacca ggtgctggat 1140
gtacagagat acccattgtt tacccaaatc acagtggaca tcgggacacc gagctag 1197

<210> 2
<211> 398
<212> PRT
<213> Homo sapiens

<400> 2

Met Arg Leu Arg Glu Pro Leu Leu Ser Gly Ser Ala Ala Met Pro Gly
 1 5 10 15

Ala Ser Leu Gln Arg Ala Cys Arg Leu Leu Val Ala Val Cys Ala Leu
 20 25 30

His Leu Gly Val Thr Leu Val Tyr Tyr Leu Ala Gly Arg Asp Leu Ser
 35 40 45

Arg Leu Pro Gln Leu Val Gly Val Ser Thr Pro Leu Gln Gly Gly Ser
 50 55 60

Asn Ser Ala Ala Ala Ile Gly Gln Ser Ser Gly Glu Leu Arg Thr Gly

2/38

65	70	75	80
Gly Ala Arg Pro Pro Pro Pro Leu Gly	Ala Ser Ser Gln Pro Arg Pro		
85	90	95	
Gly Gly Asp Ser Ser Pro Val Val	Asp Ser Gly Pro Gly Pro Ala Ser		
100	105	110	
Asn Leu Thr Ser Val Pro Val Pro His	Thr Thr Ala Leu Ser Leu Pro		
115	120	125	
Ala Cys Pro Glu Glu Ser Pro Leu Leu Val	Gly Pro Met Leu Ile Glu		
130	135	140	
Phe Asn Met Pro Val Asp Leu Glu Leu Val	Ala Lys Gln Asn Pro Asn		
145	150	155	160
Val Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp	Cys Val Ser Pro His		
165	170	175	
Lys Val Ala Ile Ile Ile Pro Phe Arg Asn Arg	Gln Glu His Leu Lys		
180	185	190	
Tyr Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg	Gln Gln Leu Asp		
195	200	205	
Tyr Gly Ile Tyr Val Ile Asn Gln Ala Gly Asp	Thr Ile Phe Asn Arg		
210	215	220	
Ala Lys Leu Leu Asn Val Gly Phe Gln Glu	Ala Leu Lys Asp Tyr Asp		
225	230	235	240
Tyr Thr Cys Phe Val Phe Ser Asp Val Asp	Leu Ile Pro Met Asn Asp		
245	250	255	
His Asn Ala Tyr Arg Cys Phe Ser Gln Pro Arg	His Ile Ser Val Ala		
260	265	270	
Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val	Gln Tyr Phe Gly Gly		
275	280	285	
Val Ser Ala Leu Ser Lys Gln Gln Phe Leu Thr	Ile Asn Gly Phe Pro		
290	295	300	
Asn Asn Tyr Trp Gly Trp Gly Glu Asp Asp Asp	Ile Phe Asn Arg		
305	310	315	320
Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn	Ala Val Val Gly		
325	330	335	
Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Lys	Asn Glu Pro Asn		
340	345	350	
Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys	Glu Thr Met Leu Ser		
355	360	365	
Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp	Val Gln Arg Tyr		
370	375	380	
Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly	Thr Pro Ser		
385	390	395	

<210> 3
<211> 1152
<212> DNA
<213> hybrid

<400> 3

atgagtaaac	ggaatcccaa	gattctgaag	attttctgt	atatgttact	tctcaactct	60
ctctttctca	tcatctactt	cgttttcac	tcatcgtcg	tttcaccgga	gcagtacag	120
cctccctata	tataccacgt	ttcagtgaat	aaccaatcg	cgatcgccgca	gtcctccggg	180
gagctccgga	ccggaggggc	ccggccgccc	cctcccttag	gcccctcc	ccagccgcgc	240
ccgggtggcg	actccagccc	agtctggat	tctggccctg	gccccgctag	caacttgacc	300
tcgggtccccag	tgcacac	caccgcactg	tcgctgccc	ctgcccctga	ggagtccccg	360
ctgcttgtgg	gccccatgct	gattgagtt	aacatgcctg	tggaccttgg	gctcgtggca	420
aaggcagaacc	caaatgtaa	gatgggcggc	cgctatgccc	ccagggactg	cgctctccct	480
cacaagggtgg	ccatcatcat	tccattccgc	aaccggcagg	agcacctcaa	gtactggcta	540
tattatttgc	acccagtccct	gcagcgcac	cagctggact	atggcatcta	tgttatcaac	600
caggcgggag	acactat	aatctgtgc	aagctctca	atgttggctt	tcaagaagcc	660
ttgaaggact	atgactacac	ctgcttgg	tttagtgacg	tggacctcat	tccaatgaat	720
gaccatgg	cgtacagg	ttttcacag	ccacggcaca	tttccgttgc	aatggataag	780
tttgattca	gcctaccta	tgttca	tttggaggt	tctctgtct	aagtaaacaa	840
cagtttctaa	ccatcaatgg	atttccata	aattattggg	gctggggagg	agaagatgat	900
gacattttta	acagattag	tttagaggc	atgtctat	ctgc	tgctgtggc	960
gggaggtgtc	gcatgatccg	ccactcaaga	gacaagaaaa	atgaacccaa	tcctcagagg	1020
tttgaccgaa	ttgcacac	aaaggagaca	atgctctgt	atggtttgaa	ctcactcacc	1080
taccaggtgc	tggatgtaca	gagataccca	ttgtatacc	aaatcacagt	ggacatcg	1140
acaccgagct	ag					1152

<210> 4

<211> 383

<212> PRT

<213> hybrid

<400> 4

Met	Ser	Lys	Arg	Asn	Pro	Lys	Ile	Leu	Lys	Ile	Phe	Leu	Tyr	Met	Leu
1					5			10					15		

Leu	Leu	Asn	Ser	Leu	Phe	Leu	Ile	Ile	Tyr	Phe	Val	Phe	His	Ser	Ser
		20						25					30		

Ser	Phe	Ser	Pro	Glu	Gln	Ser	Gln	Pro	Pro	His	Ile	Tyr	His	Val	Ser
	35					40				45					

Val	Asn	Asn	Gln	Ser	Ala	Ile	Gly	Gln	Ser	Ser	Gly	Glu	Leu	Arg	Thr
	50					55				60					

Gly	Gly	Ala	Arg	Pro	Pro	Pro	Leu	Gly	Ala	Ser	Ser	Gln	Pro	Arg	
b5					70				75			80			

Pro	Gly	Gly	Asp	Ser	Ser	Pro	Val	Val	Asp	Ser	Gly	Pro	Gly	Pro	Ala
	85					90			95						

Ser	Asn	Leu	Thr	Ser	Val	Pro	Val	Pro	His	Thr	Thr	Ala	Leu	Ser	Leu
	100					105				110					

Pro	Ala	Cys	Pro	Glu	Glu	Ser	Pro	Leu	Leu	Val	Gly	Pro	Met	Leu	Ile
	115					120				125					

Glu	Phe	Asn	Met	Pro	Val	Asp	Leu	Glu	Leu	Val	Ala	Lys	Gln	Asn	Pro
	130					135				140					

Asn	Val	Lys	Met	Gly	Gly	Arg	Tyr	Ala	Pro	Arg	Asp	Cys	Val	Ser	Pro
	145					150				155			160		

His	Lys	Val	Ala	Ile	Ile	Ile	Pro	Phe	Arg	Asn	Arg	Gln	Glu	His	Leu
	165						170		175						

Lys	Tyr	Trp	Leu	Tyr	Tyr	Leu	His	Pro	Val	Leu	Gln	Arg	Gln	Gln	Leu
	180						185			190					

Asp	Tyr	Gly	Ile	Tyr	Val	Ile	Asn	Gln	Ala	Gly	Asp	Thr	Ile	Phe	Asn
	195						200			205					

Arg Ala Lys Leu Leu Asn Val Gly Phe Gln Glu Ala Leu Lys Asp Tyr
 210 215 220
 Asp Tyr Thr Cys Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met Asn
 225 230 235 240
 Asp His Asn Ala Tyr Arg Cys Phe Ser Gln Pro Arg His Ile Ser Val
 245 250 255
 Ala Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val Gln Tyr Phe Gly
 260 265 270
 Gly Val Ser Ala Leu Ser Lys Gln Gln Phe Leu Thr Ile Asn Gly Phe
 275 280 285
 Pro Asn Asn Tyr Trp Gly Trp Gly Glu Asp Asp Asp Ile Phe Asn
 290 295 300
 Arg Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val
 305 310 315 320
 Gly Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Lys Asn Glu Pro
 325 330 335
 Asn Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met Leu
 340 345 350
 Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg
 355 360 365
 Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser
 370 375 380

<210> 5
 <211> 1642
 <212> DNA
 <213> Homo sapiens

<400> 5
 ccatggat gagacgctac aagctttc tcatgttctg tatggccggc ctgtgcctca 60
 ttccttcct gcacttctc aagaccctgt cctatgtcac ttccccccga gaactggcct 120
 ccctcagccc taacctggtg tccagcttt tctggAACAA tgccccggtc acgccccagg 180
 ccagccccga gccaggaggc cctgacctgc tgcgtaccccc actctactcc cactcgcccc 240
 tgctgcagcc gctgccgccc agcaaggccg ccgaggagct ccaccgggtg gacttggtgc 300
 tgcccggagga caccaccgag tatttcgtgc gcacccaaggc cggcggcgtc tgcttcaaac 360
 cccgcaccaa gatgctggag aggccgcccc cgggacggcc ggaggagaag cctgaggggg 420
 ccaacggctc ctccggcccg cggccacccc ggtacccctt gaggccccgg gagcgcacgg 480
 ggggcccggagg cgccggcgc aagtgggtgg agtgcgtgtg cctggccggc tggcacggac 540
 ccagctgcgg cgtcccaact gtggtgcagt actccaaccc gcccaccaag gagcggctgg 600
 tgcccggagga ggtccggcgc cgcgtcatca acgccccaa cgtcaaccac gagttcgacc 660
 tgctggacgt gcgttccac gagctggccg acgtgggtgg cggctttgtg gtgtgcgagt 720
 ccaacttcac ggcttatggg gagccgcggc cgctcaagtt cgggagatg ctgaccaatg 780
 gcaccttcga gtacatccgc cacaagggtgc tctatgtctt cctggaccac ttccggcccg 840
 gcccggcga ggacggctgg atcggcgcacg actacccgtcg caccttcctc acccaggacg 900
 gctgttcgcg gctgcgcac acgtggccg acgtgggtgg cggctttgtg gtgtgcgagt 960 1020
 agatccccggc cctgtacggc gtccttttcc tcaagctca cgatggctgg accgagccct 1080
 tcgccttcga catgcgcacg tgcgtctacg gtttcttcg aaaggcggcc ggcacccctgg 1140
 aggtgggtgc aggctgcacg gtggacatgc tgcaggcagt gtatggctg gacggcatcc 1200
 gcctgcggcgc cccggcgtac tacaccatgc ccaacttcac acatgtatgaa aaccgcaccc 1260
 gcccacatcc tggtcgtgg tcgtggccgac gccccctgca cttccggccgc tggcaactgt 1320
 cctggtgctt cacggccggc ggcacatctact tcaagctcggt gtcggcccgag aatggcgact 1380
 tccccacgctg gggtgactac gaggacaacg gggacctgaa ctacatccgc ggcctgtatcc 1440
 gcacccggggg ctggttcgcac ggcacgcacg aggatcaccc gcctgcagac cccagcgacg 1500
 acatgtatgc gcccacatgc ctgctgaaga actacgcaccc gttccactac ctgctggaca 1560
 aaccctacca ggagcccgagg agcacaacggc cgggggggtg ggcacccacagg ggtcccgagg 1620
 gaaggccggcc cggccggggc aaactggacg aggccggaaat ctcatctac

aagaggatct gaattaggat cc

1642

<210> b
<211> 544
<212> PRT
<213> Homo sapiens

<400> b

Met Val Met Arg Arg Tyr Lys Leu Phe Leu Met Phe Cys Met Ala Gly
1 5 10 15

Leu Cys Leu Ile Ser Phe Leu His Phe Phe Lys Thr Leu Ser Tyr Val
20 25 30

Thr Phe Pro Arg Glu Leu Ala Ser Leu Ser Pro Asn Leu Val Ser Ser
35 40 45

Phe Phe Trp Asn Asn Ala Pro Val Thr Pro Gln Ala Ser Pro Glu Pro
50 55 60

Gly Gly Pro Asp Leu Leu Arg Thr Pro Leu Tyr Ser His Ser Pro Leu
65 70 75 80

Leu Gln Pro Leu Pro Pro Ser Lys Ala Ala Glu Glu Leu His Arg Val
85 90 95

Asp Leu Val Leu Pro Glu Asp Thr Thr Glu Tyr Phe Val Arg Thr Lys
100 105 110

Ala Gly Gly Val Cys Phe Lys Pro Gly Thr Lys Met Leu Glu Arg Pro
115 120 125

Pro Pro Gly Arg Pro Glu Glu Lys Pro Glu Gly Ala Asn Gly Ser Ser
130 135 140

Ala Arg Arg Pro Pro Arg Tyr Leu Leu Ser Ala Arg Glu Arg Thr Gly
145 150 155 160

Gly Arg Gly Ala Arg Arg Lys Trp Val Glu Cys Val Cys Leu Pro Gly
165 170 175

Trp His Gly Pro Ser Cys Gly Val Pro Thr Val Val Gln Tyr Ser Asn
180 185 190

Leu Pro Thr Lys Glu Arg Leu Val Pro Arg Glu Val Pro Arg Arg Val
195 200 205

Ile Asn Ala Ile Asn Val Asn His Glu Phe Asp Leu Leu Asp Val Arg
210 215 220

Phe His Glu Leu Gly Asp Val Val Asp Ala Phe Val Val Cys Glu Ser
225 230 235 240

Asn Phe Thr Ala Tyr Gly Glu Pro Arg Pro Leu Lys Phe Arg Glu Met
245 250 255

Leu Thr Asn Gly Thr Phe Glu Tyr Ile Arg His Lys Val Leu Tyr Val
260 265 270

Phe Leu Asp His Phe Pro Pro Gly Gly Arg Gln Asp Gly Trp Ile Ala
275 280 285

Asp Asp Tyr Leu Arg Thr Phe Leu Thr Gln Asp Gly Val Ser Arg Leu
290 295 300

Arg Asn Leu Arg Pro Asp Asp Val Phe Ile Ile Asp Asp Ala Asp Glu

6/38

305	310	315	320
Ile Pro Ala Arg Asp Gly Val Leu Phe Leu Lys Leu Tyr Asp Gly Trp			
325		330	335
Thr Glu Pro Phe Ala Phe His Met Arg Lys Ser Leu Tyr Gly Phe Phe			
340		345	350
Trp Lys Gln Pro Gly Thr Leu Glu Val Val Ser Gly Cys Thr Val Asp			
355		360	365
Met Leu Gln Ala Val Tyr Gly Leu Asp Gly Ile Arg Leu Arg Arg Arg			
370	375		380
Gln Tyr Tyr Thr Met Pro Asn Phe Arg Gln Tyr Glu Asn Arg Thr Gly			
385	390	395	400
His Ile Leu Val Gln Trp Ser Leu Gly Ser Pro Leu His Phe Ala Gly			
405		410	415
Trp His Cys Ser Trp Cys Phe Thr Pro Glu Gly Ile Tyr Phe Lys Leu			
420		425	430
Val Ser Ala Gln Asn Gly Asp Phe Pro Arg Trp Gly Asp Tyr Glu Asp			
435		440	445
Lys Arg Asp Leu Asn Tyr Ile Arg Gly Leu Ile Arg Thr Gly Gly Trp			
450		455	460
Phe Asp Gly Thr Gln Gln Glu Tyr Pro Pro Ala Asp Pro Ser Glu His			
465		470	475
Met Tyr Ala Pro Lys Tyr Leu Leu Lys Asn Tyr Asp Arg Phe His Tyr			
485		490	495
Leu Leu Asp Asn Pro Tyr Gln Glu Pro Arg Ser Thr Ala Ala Gly Gly			
500		505	510
Trp Arg His Arg Gly Pro Glu Gly Arg Pro Pro Ala Arg Gly Lys Leu			
515		520	525
Asp Glu Ala Glu Val Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Asn			
530		535	540

<210> ?
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> ?

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 8
<211> 31
<212> PRT
<213> Homo sapiens

<400> 8

Gln Glu Pro Arg Ser Thr Ala Ala Gly Gly Trp Arg His Arg Gly Pro
1 5 10 15

Glu Gly Arg Pro Pro Ala Arg Gly Lys Leu Asp Glu Ala Glu Val
 20 25 30

<210> 9
 <211> 1614
 <212> DNA
 <213> hybrid

<400> 9
 catgagtaaa cggaatccga agattctgaa gatTTTCTG tATATGTTAC ttCTCAACTC
 tCTCTTCTC atCATCTACt tcgttttca ctcATGTcG ttttCACCGG agcAGTCACA
 gcCTCCTCAT atataccacG tttcAGTgAA taACCAATCG gcACATGGAG gCCCTGACCT
 gCTGCgtTACc ccACTCTACT cCCACTCGCC cCTGCTGCGAG CCAGCTGCCG
 ggCCGAGGAG ctCCACCGGG tggacttggT gCTGCCGAG gACACCACCG agTATTGCT
 gCGCACCAAG gCCGGCGCGC tCTGCTTCAA ACCCGCACC aAGATGCTGG agAGGCCGCC
 CCCGGGACGG CGGGAGGAGA AGCCTGAGGG gGCCAACCGC tcCTCGGCCG gGCggCCACC
 CCGGTACCTC CTGAGCGCCG gggAGCGCAC gggggGCCGA ggcGCCCGGC gCAAGTGGGT
 ggAGTGCgtG tgCCtGCCG gCTGGCACGG ACCCAGCTGC ggcGtGCCCA CTGTTGtGCA
 gtACTCCAAC CTGCCCACCA AGGAGCAGGt ggtGCCAGG gagGTGCCGC gCCGCGTcat
 caACGCCATC AACGTCAACC AGCAGTTCGA CCTGCTGGAC gTGCgtTCC acGAGCTGG
 CGACGtGGTG gACGCTTtG tggTGTGcGA gTCCAACTTC acGGCTTATG gggAGCCGCG
 gCCGCTCAAG ttCCGGGAGA tgCTGACCAA tggCACCTC gagTACATCC gCCACAAAGT
 gCTCTATGTC ttCCtGGACG ACTTCCCGCC CGGCGGCCGG CAGGACGGCT ggATGCCGA
 CGACTACCTG CGCACCTTCC TCAACCCAGGA CGGCGTCTG CGGCTGCGCA ACCTGCCGCC
 CGACGACGTC ttCATCATTG ACgATGCGGA CGAGATCCCG GCCCgtGACG gCgtCCttt
 CCTCAAGTC tacGATGGCT ggaccGAGCC CCTCGCCtC CACATGCGCA AGTCGCTCTA
 CGGCTTCTC tGGAAGCAGC CGGGCACCTC ggAGGTGGTG TCAAGGCTGCA CGGTGGACAT
 gCTGCAgGCA GTGTATGGGC tggACGGCAT CGGCGTGCgC CGGCCGCAgT actACACCAT
 gCCCAACTTC AGACAGTATG AGAACCGCAC CGGCCACATC CTGGTGCAGT ggtCGtGGG
 CAGCCCCCTG CACTCGCCG GCTGGACTG CCTCTGGTG tTCACGCCG aggGcatCTA
 CCTCAAGTC GTGTCCGCCC AGAATGGCGA CTTCCACGC TGGGGTgACT acGAGGACAA
 CGGGGACCTG AACTACATCC CGGGCCTGAT CGGCAACCGGG ggCTGGTTCG acGGCACGCA
 GCAGGAGTAC CGGCGTGCAG ACCCCAGCGA GCACATGTAT GCGCCCAAGT ACCTGCTGAA
 GAACtACGAC CGGTCCACT ACCTGCTGGA CAACCCCTAC CAGGAGCCCA ggAGCACGGC
 ggCGGGCGGG tggGCCACA gggGTCCCGA gggAAAGGCCG CCCGCCCGGG gCAAActGGa
 CGAGGCGGAA gTCGAACAAA AACTCATCTC AGAAGAGGAT CTGAATTAGG ATCC 1614

<210> 10
 <211> 535
 <212> PRT
 <213> hybrid

<400> 10

Met Ser Lys Arg Asn Pro Lys Ile Leu Lys Ile Phe Leu Tyr Met Leu
 1 5 10 15

Leu Leu Asn Ser Leu Phe Leu Ile Ile Tyr Phe Val Phe His Ser Ser
 20 25 30

Ser Phe Ser Pro Glu Gln Ser Gln Pro Pro His Ile Tyr His Val Ser
 35 40 45

Val Asn Asn Gln Ser Ala His Gly Gly Pro Asp Leu Leu Arg Thr Pro
 50 55 60

Leu Tyr Ser His Ser Pro Leu Leu Gln Pro Leu Pro Pro Ser Lys Ala
 65 70 75 80

Ala Glu Glu Leu His Arg Val Asp Leu Val Leu Pro Glu Asp Thr Thr
 85 90 95

Glu Tyr Phe Val Arg Thr Lys Ala Gly Gly Val Cys Phe Lys Pro Gly
 100 105 110

8/38

Thr Lys Met Leu Glu Arg Pro Pro Pro Gly Arg Pro Glu Glu Lys Pro
115 120 125

Glu Gly Ala Asn Gly Ser Ser Ala Arg Arg Pro Pro Arg Tyr Leu Leu
130 135 140

Ser Ala Arg Glu Arg Thr Gly Gly Arg Gly Ala Arg Arg Lys Trp Val
145 150 155 160

Glu Cys Val Cys Leu Pro Gly Trp His Gly Pro Ser Cys Gly Val Pro
165 170 175

Thr Val Val Gln Tyr Ser Asn Leu Pro Thr Lys Glu Arg Leu Val Pro
180 185 190

Arg Glu Val Pro Arg Arg Val Ile Asn Ala Ile Asn Val Asn His Glu
195 200 205

Phe Asp Leu Leu Asp Val Arg Phe His Glu Leu Gly Asp Val Val Asp
210 215 220

Ala Phe Val Val Cys Glu Ser Asn Phe Thr Ala Tyr Gly Glu Pro Arg
225 230 235 240

Pro Leu Lys Phe Arg Glu Met Leu Thr Asn Gly Thr Phe Glu Tyr Ile
245 250 255

Arg His Lys Val Leu Tyr Val Phe Leu Asp His Phe Pro Pro Gly Gly
260 265 270

Arg Gln Asp Gly Trp Ile Ala Asp Asp Tyr Leu Arg Thr Phe Leu Thr
275 280 285

Gln Asp Gly Val Ser Arg Leu Arg Asn Leu Arg Pro Asp Asp Val Phe
290 295 300

Ile Ile Asp Asp Ala Asp Glu Ile Pro Ala Arg Asp Gly Val Leu Phe
305 310 315 320

Leu Lys Leu Tyr Asp Gly Trp Thr Glu Pro Phe Ala Phe His Met Arg
325 330 335

Lys Ser Leu Tyr Gly Phe Phe Trp Lys Gln Pro Gly Thr Leu Glu Val
340 345 350

Val Ser Gly Cys Thr Val Asp Met Leu Gln Ala Val Tyr Gly Leu Asp
355 360 365

Gly Ile Arg Leu Arg Arg Gln Tyr Tyr Thr Met Pro Asn Phe Arg
370 375 380

Gln Tyr Glu Asn Arg Thr Gly His Ile Leu Val Gln Trp Ser Leu Gly
385 390 395 400

Ser Pro Leu His Phe Ala Gly Trp His Cys Ser Trp Cys Phe Thr Pro
405 410 415

Glu Gly Ile Tyr Phe Lys Leu Val Ser Ala Gln Asn Gly Asp Phe Pro
420 425 430

Arg Trp Gly Asp Tyr Glu Asp Lys Arg Asp Leu Asn Tyr Ile Arg Gly
435 440 445

Leu Ile Arg Thr Gly Gly Trp Phe Asp Gly Thr Gln Gln Glu Tyr Pro
450 455 460

Pro Ala Asp Pro Ser Glu His Met Tyr Ala Pro Lys Tyr Leu Leu Lys
465 470 475 480

Asn Tyr Asp Arg Phe His Tyr Leu Leu Asp Asn Pro Tyr Gln Glu Pro
 485 490 495
 Arg Ser Thr Ala Ala Gly Gly Trp Arg His Arg Gly Pro Glu Gly Arg
 500 505 510
 Pro Pro Ala Arg Gly Lys Leu Asp Glu Ala Glu Val Glu Gln Lys Leu
 515 520 525
 Ile Ser Glu Glu Asp Leu Asn
 530 535

<210>	11			
<211>	13			
<212>	DNA			
<213>	Artificial Sequence			
<220>				
<223>	Synthetic			
<400>	11			
aataacttccaa	ccc	13		
<210>	12			
<211>	34			
<212>	DNA			
<213>	Artificial Sequence			
<220>				
<223>	Synthetic			
<400>	12			
ccacccgtta	acaatgaaga	tgagacgcta	caag	34
<210>	13			
<211>	29			
<212>	DNA			
<213>	Artificial Sequence			
<220>				
<223>	Synthetic			
<400>	13			
gggccatgga	gatgagacgc	tacaagctc		29
<210>	14			
<211>	28			
<212>	DNA			
<213>	Artificial Sequence			
<220>				
<223>	Synthetic			
<400>	14			
ggatccaaatg	aagatgagac	gctacaag		28
<210>	15			
<211>	41			
<212>	DNA			
<213>	Artificial Sequence			
<220>				

10/38

<223> Synthetic

<400> 15
gggccccggga gatcctaatt cagatcctct tctgagatga g 41<210> 16
<211> 35
<212> DNA
<213> Artificial Sequence<220>
<223> Synthetic<400> 16
cccgaggatcct aattcagatc ctcttctgag atgag 35<210> 17
<211> 38
<212> DNA
<213> Artificial Sequence<220>
<223> Synthetic<400> 17
gggtcttagat cctaattcag atcctttct gagatgag 38<210> 18
<211> 35
<212> DNA
<213> Artificial Sequence<220>
<223> Synthetic<400> 18
ccaccccggtta acaatgagta aacggaatcc gaaga 35<210> 19
<211> 35
<212> DNA
<213> Artificial Sequence<220>
<223> Synthetic<400> 19
gggccatggg taaacggaat ccgaagattc tgaag 35<210> 20
<211> 34
<212> DNA
<213> Artificial Sequence<220>
<223> Synthetic<400> 20
cccgaggatcca tgagttaaacg gaatccgaag attc 34<210> 21
<211> 29

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 21
gcgccccggg acgctagctc ggtgtcccg 29

<210> 22
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 22
cccggatcca cgctagctcg gtgtc 25

<210> 23
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 23
gggtcttagat ccacgctagc tcgggtgtccc g 31

<210> 24
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 24
ccacccgtta acaatgaggc ttcgggagcc gtcctgag 39

<210> 25
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 25
gggccatggg gcttcgggag ccgctcctga g 31

<210> 26
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 26
cccgatcca tgaggcttcg ggagccgctc ctgag 35

<210> 27
<211> 7155
<212> DNA
<213> hybrid

<400> 27						
ggcgcgcctc	gaggcgatcg	cagatcta	at ctaaccatt	acgatacgct	ttgggtacac	60
ttgatTTTt	tttcaGtggt	tacatata	ttgtttata	tgttatctt	aaggatctgc	120
acaaaagatta	tttGttgtat	ttcttgatgg	ggctcagaag	atttgatatg	atacactcta	180
atcttttagga	gataccagcc	aggattata	tcaGtaagac	aatcaaattt	tacgtgttca	240
aactcgttat	cttttcatc	aaaggatgag	ccagaatctt	tatagaatga	ttgcaatcga	300
gaatatgttc	ggccgatatg	ccttGtttg	cttcaatatt	cttcatatca	cacaagaatc	360
gaccgtattt	taccCtCttt	ccataaaaggaa	aaacacaata	tgcagatgct	ttttcccac	420
atgcagtaac	ataatggat	tcaaaaaatgg	ctaaaagaag	ttggataaca	aattgacaac	480
tatTTccatt	tctgttat	aaatTCaca	acacacaaaa	gcccgtatc	aagagtctgc	540
ccatgtacga	aataacttct	attatTTgtt	attggccta	agcccagctc	agagtacgtg	600
ggggtaccac	atataggaa	gtaacaaaat	actgcaagat	agccccataa	cgtaccagcc	660
tctccttacc	acgaagagat	aagatataag	acccacccctg	ccacgtgtca	catcgctatg	720
gtggtaatg	ataagggtt	acatCtttct	atgtttgtgg	acatgatgca	tgtaatgtca	780
tgagccacag	gatccaatgg	ccacaggaac	gtaagaatgt	agatagattt	gattttgtcc	840
gttagatagc	aaacaacatt	ataaaaagggt	tgtatcaata	ggaacttaatt	cactcattgg	900
attcatagaa	gtccatttct	cttaaGgtatc	tagaaacat	ggcgaggatc	tcgtgtgact	960
tgagatttt	tctcatccc	gcagcttca	tgttcatcta	catccagatg	aggctttcc	1020
agacgcatac	acagatatgca	gatcgctca	gttccgtat	cgaatctgag	aaccatcgca	1080
ctagtcaat	gcgaggcctc	atagatgaag	ttagcatcaa	acagtcgcgg	attgttccc	1140
tcgaagatat	gaagaaccgc	caggacgaag	aacttGtca	gcttaaggat	ctaattccaga	1200
cgTTTgaaaa	aaaaggaa	gcaaaactca	ctcaagggtgg	agccatggat	tccaattcag	1260
gcgcgtcgt	tgatatcaca	actaaagatc	tatacgatag	gattgaggat	cttgatacag	1320
atgggtgtcc	atggaaacaa	gttggagag	ttacgtataa	agacgtatgg	tgggagaaag	1380
agaagctcaa	aatcccggt	gttccattt	ctcataacga	tcttgggtgg	aaatttgcgt	1440
tagaggagta	ttatcagaga	caatccagac	atatttttga	caccattgtt	gagactttat	1500
ctaaggatg	acgaaagttt	ttgcttttgg	ttttatattt	ttaattctct	cccatgggta	1560
tcccgtgaac	aatcttaaat	gtcttaaat	tctcatgacg	tcattaaact	ctataaccaa	1620
acttcttgc	tgggttctgt	tttttttag	ttcgtgtat	aaacagagtt	ctagaagttc	1680
gttctttgg	aaaatttggaa	gtctttggag	ctaaagttt	ttttttattt	actgggtttt	1740
gagattgaag	gatagctaga	atcttattt	tgtgggggtt	tgttttgaat	atgttttata	1800
ggattcaaga	agaaaagtta	tatgggagga	gatgtcatat	ctggagagat	ggtggagaga	1860
cgcttccacct	aataaacaag	aagctttgac	taaattgggtt	aggatgggc	agctagagat	1920
tgttggaggt	ggctgggtta	tgaatgtat	ggctattca	cattattttt	ccataattga	1980
acagatagca	gagggtaata	tgtggctgaa	tgacacaatt	gggggttattc	ctaagaattc	2040
ttgggctata	gatcccttgc	gttattatc	aaccatggct	tatcttctcc	ggcgtatggg	2100
tttgaaaac	atgttattt	aaaggactca	ttacgagctc	agaaaagacc	ttgcccagca	2160
taagaatctt	gaatataattt	ggcgtcagag	ctgggatgt	atggaaacca	cagatatctt	2220
tgttcatatg	atgcccgtttt	attcatacga	tatcccacac	acttGtggac	cagagccgtc	2280
aatttgcgt	cagtttgatt	tcgctcgat	gcggggattt	aagtatgaa	tttgtccatg	2340
ggggaaagcac	ccagtggaga	ccacactaga	aaatgtgcag	gagaggggat	taaagcttct	2400
ggatcaatac	aggaaaaaaat	ccactctata	tcgactaat	acacttctta	taccttttgg	2460
agatgatTTT	aggttacat	gtatcgat	agccgaggat	cagttccgt	actaccagat	2520
gttGtttGat	cacatcaact	ctaatcctag	tctaaacgc	gaaGcaaaat	tttgtacttt	2580
ggaggatttt	ttcagaacag	tccgagaaga	agcagacaga	gtgaattatt	ctcgctctgg	2640
tgggttggc	tctggtcagg	ttgttgggtt	cccttctctg	tcaGgtgtact	tctttacata	2700
tgcagatagg	caacaagact	attggatgtt	ttattatgtt	tcaagacctt	tcttcaaagc	2760
tgttGatcg	gtgtcgagc	atacccttc	tggagctgag	atcatgatgt	catttGtct	2820
aggttattgc	catcgaaatc	aatgtgagaa	atTTcaaca	agttttacgt	ataagttgac	2880
tgcgtcaaga	agaaatctgg	ctttttcca	gcaccatgtat	ggggtaactg	gaactgtctaa	2940
ggattatgt	gtacaagatt	acggcaccgc	gatgcatact	tcattGtcaag	accttcagat	3000
ctttatgtct	aaAGcaatcg	aagtctttt	tggatccgc	cacgagaaag	aaaaatctga	3060
tcaatcccc	tcattttcg	aggcagagac	aatgagatca	aagtatgatg	ctcgcccagt	3120
tcaacaagca	attgtgccc	gggaaaggaaa	ttcgcacaca	gttatactct	tcaattccatc	3180
agaacagacg	agagaggagg	tggtgacggt	tgttGtttaac	cgcgctgaaa	tctcggttt	3240
ggactcaaac	tggacttgc	tccctagcca	aatttctctt	gaagtgcagc	atgacgatac	3300
caaactattc	acccggcagac	atcgccTTA	ctggaaagct	tccatcccag	ctcttggtct	3360
gagaacatata	ttcatttgct	atgggaatgt	cgagtgtgag	aaagctactc	cgtctaaact	3420
caaatacgct	tctgagtttgc	accatTTCC	ttgttccct	ccatattctt	gctccaaact	3480
ggacaacgac	gttactgaga	tccgaaatga	acatcgact	tttGtttgg	atgtgaagaa	3540
cggtactcgt	cggaagatag	tccatagaaa	cggtatcgag	acttGtttgg	gagaagagat	3600

aggtatgtac tctagtcag agagtggagc ttacctgttc aaaccagatg gtgaagctca	3660
gccaattgtt caacccgtat gacatgtat cacctctgag ggtctgtcg ttcaagaagt	3720
cttctttac cctaaaacca aatggggagaa atcacccctc tctcagaaaaa ctgcgtctta	3780
cactggaggt aatacgttc aggatcaagt ggtcgagata gaatatcatg ttgagcttct	3840
tggtaatgtt tttgtatgacc gggaaatgtat tgccggatc aagactgtat ttgacaacaa	3900
gaagggtcttc tatttagatc tcaatggttt ccaaatacgac aggagagaaaa cttatgataa	3960
gatccctttt caaggaaact actacccaaat gccatctctc gcatttatcc aaggatccaa	4020
tggtcagaga ttctccgtgc actctcgta atctctcggt ttgcagaccc tcaaagaggg	4080
ttgggtggag attatgtctgg acagacgggtt gggtcgatc gacggacggg gtcttagggca	4140
agggtgtatg gataaccgcg caatgaccgt ggtatttcac ctcttgcgg aatctaataat	4200
ttctcaagca gaccctgtt ccaacactaa cccgagggac ctttcgttcc tctctcacct	4260
cataggtgtt cactaaact accccataaa cacatttattt gccaagaaaac cgcaagacat	4320
atctgtgtgtt gttccacaaat acggttccct tgctccctta gccaaacccgt taccatgtga	4380
cctccacattt gtaaatttttcc aggttccctcg tccatccaaa tactctcagc aattggaaaga	4440
agacaaggcca aggttcgttc ttatcctcaat tagacgagct tgggatttcag cttatttcca	4500
taaaggaaaga caagtaaact gcacaaggat ggctaataat ccagtaaact tttccgacat	4560
gttcaaagat cttgcagctt caaaggtaaa accaacttca ctgaaatctct tgcaagaaga	4620
tatggagatt ttgggtatc atgaccaaga gctacccgt gatagtttac agccacggg	4680
aggacgtgtc tcgatctctc ccatggaaat acgagttat aagcttgcac tgcgacttca	4740
caagtgaacc tgctgaagat ccgcttagat ccgcaaaaat caccaggctc tctctacaaa	4800
tctatctctc tctatttttc tccagaataat tgggtgtatc ttcccagat aaggaaatta	4860
gggttcttat aggggttcgc tcatgtgttg agcatatgg aaacccttag tatgtatgg	4920
tatttgtaaa atatcttat caataaaaat tctaataatccaa aacccaaaat cccgcgcgcg	4980
cctcgaggcg atcgacatc tcattatacc gttagaagca tagttaaaaat ctaaagcttg	5040
tcgttaattt tagtcatatccatcattttt acattgttgg gttctacattt attaatgaaat tttctaataatgc	5100
aaatacagaa tttaatcaatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataatct	5160
ccgccttgcg tgggtgtatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5220
ctatgagaag gcagggtggca acacaaaacaa gagtatctaa gattttcat tggactata	5280
ggaatataat atcttttacatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5340
aagatcacaa ctatgtatccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5400
ttcaacttagc cccacaaaat acattgttccc ttatgttgc accttttgtt tttatattat	5460
tcttggggat ctaatgttgc atatcttatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5520
gagaagaaaa ccatggcgat gatctcgat gacttggat ttcttctcat cccggcagct	5580
ttcatgttca tctatcatccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5640
ctcagttccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5700
gaagtttagca tcaaacatgc gcgattttt gcccctcgaaat gatgtatcaat catgtacat acgtataat	5760
gaagaacttgc tcaatgttgc atatcttatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5820
ctcaactcaatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5880
aatacggatc tgggtttccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	5940
aatcgggtctc agtattttcg agtacatgc gaaagggttgc atatgtatcaat catgtacat acgtataat	6000
gagacatgt tgggttttag tcatgtatgc tttatgttgc atatgtatcaat catgtacat acgtataat	6060
agtattttatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6120
actatgttccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6180
cattgttgcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6240
aagcatcaatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6300
catgaggggc atatctttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6360
aacatacaga ctcttacatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6420
tttagcaccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6480
atggggaaatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6540
agagagttttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6600
ccgtcgatgttgc tttatgttggaa ttatgtatcaat catgtacat acgtataat	6660
ggaaaatgttgc tttatgttggaa ttatgtatcaat catgtacat acgtataat	6720
gtttatgttgc tttatgttggaa ttatgtatcaat catgtacat acgtataat	6780
cggtgtatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6840
gatgtatggat accgacatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6900
agtgcacatccatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	6960
aatctatctc tcttcttacatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	7020
taggggtttt acattgttggaa ttatgtatcaat catgtacat acgtataat	7080
tgtatgttgc tttatgttggaa ttatgtatcaat catgtacat acgtataat	7140
agaccttttacatcattttt acattgttggaa ttatgtatcaat catgtacat acgtataat	7155

<210> 28
<211> 4993
<212> DNA
<213> hybrid

<400> 28

ggcgcgcctc	gaggcgatcg	cagatctaatt	ctaaccattt	acgatacgctt	ttgggtacac	60
ttgatTTTt	tttcagtgg	tacatatactt	ttgtttata	tgcstatctt	aaggatctgc	120
acaaagatta	tttgttgatg	ttcttgatgg	ggctcagaag	attgtatgt	atacactctt	180
atcttttagga	gataccagcc	aggatttat	tcaagatgg	aatcaaattt	tacgtgttca	240
aactcggtt	cttttcatc	aaaggatgg	ccagaatctt	tatagaatga	ttgcaatcga	300
aatatgttc	ggccgatat	cTTTGTG	cttcaatatt	ctacatatca	cacaagaatc	360
gaccgtattt	taccccttt	ccataaaggaa	aaacacaata	tgagatgtct	tttttcccac	420
atgcagtaac	atataggat	tcaaaaatgg	ctaaaagaag	ttggataaca	aattgacaac	480
tatttccatt	tctgttat	aaatttcaca	acacacaaaa	gcccgtatc	aagagtctgc	540
ccatgtacga	aataacttct	attatttgg	attgggccta	agccccagctc	agagtacgtg	600
ggggtaccac	ata>taggaag	gtaacaaaat	actgcaagat	agccccataa	cgtaccagcc	660
tctccattacc	acgaagagat	agatataaag	accacccctg	ccacgtgtca	catcgatcg	720
gtggtaatg	ataagggatt	acatccctt	atgttgttgg	acatgtgc	tgtaatgtca	780
tgagccacag	gatccaatgg	ccacaggaac	gtaagaatgt	agatagattt	gatttgtcc	840
gttagatagc	aaacaaacattt	ataaaaagggt	tgtataataa	ggaactaattt	cactcattgg	900
attcatagaa	gtcccttctt	ctaaagtatc	tagaaaccat	ggcggaggatc	tcgtgtact	960
tgagatttct	tctcatcccg	gcagcttca	tgttcatcta	catccagatg	aggctttcc	1020
agacgcaatc	acagatgtca	gatcgctca	gttccgtat	cgaatctgag	aaccattgca	1080
ctagtcaat	gcgaggcctc	atagatgaag	ttagcatca	acagtcgcgg	attgttccc	1140
tcgaagatat	gaagaaccgc	caggacgaag	aacttgtca	gcttaaggat	ctaattccaga	1200
cgtttaaaaa	aaaaggaata	gcaaaactca	ctcaaggtgg	accatggat	tccaatccag	1260
gcgcgcgtgt	tgatatcaca	actaaagatc	tatacgtat	gattgagttt	tttgatacag	1320
atgggtgtcc	atggaaacaa	gttgggagag	ttacgtataa	agacgtatg	tgggagaaag	1380
agaagtcata	aatcccggtt	ttcccttatt	ctcataacga	tccttgttgg	aaattgactg	1440
tagaggagat	ttatccagaga	caatccagac	atattttca	caccattgtt	gagactttat	1500
ctaaggatgt	acgaaagttt	ttgtttttgg	ttttatattt	ttaattctct	cccatgttta	1560
tcccgtgaac	aatcttaaat	gtcttaaat	tctcatgacg	tcattaaact	ctataaccaa	1620
acttcttgc	tgggttctgt	tttttttag	tttcgtgtat	aaacagagtt	ctagaagttc	1680
gttcttttgg	aaaatttggaa	gtctttggag	ctaaagttt	tttttttatt	actgggtttt	1740
gagattgaag	gatagtctaga	atcttattt	tgtgggggtt	tgttttgaat	atgtttataa	1800
ggattcaaga	agaaaagttt	tatgggagga	gatgtcatat	ctggagatgt	ggtggagaga	1860
cgcttcacct	aataaacaag	aagctttgac	taaatgtgg	aaggatgggc	agcttagat	1920
tgttggaggt	ggctggggtt	tgaatgtat	ggctaatttca	cattattttt	ccataatttga	1980
acagatgtca	gagggttataa	tgtggctgaa	tgacacaaattt	ggggttattt	ctaagaattt	2040
ttgggctata	gatccctttt	gcttatttcatc	aaccatggct	tatcttctcc	ggcgtatggg	2100
ttttgaaaac	atgtttattt	aaaggactca	ttacgagctc	aagaaagacc	ttgcccagca	2160
taagaatctt	gaatatattt	ggcgtcagag	ctgggatgt	atggaaacca	cagatattttt	2220
tgttcatatg	atgcgtttt	attcatacga	tatcccacac	acttgtggac	cagagcctgc	2280
aatttgcgt	cagtttgatt	tcgctcgat	gcggggattt	aagtatgttca	tttgtccatg	2340
ggggaaagcac	ccagtggtt	ccacactaga	aaatgtgcag	gagaggggat	taaagtttct	2400
ggatcaatac	aggaaaaaat	ccactctata	tcgaactaat	acacttcttta	taccttttgg	2460
agatgatttt	aggtacattt	gtatcgatga	agccgaggct	caagtccgtt	actaccagat	2520
ttgttggat	cacatcaact	ctaattctcg	tctaaacgc	gaagcaaaat	tttgtacttt	2580
ggaggatttt	ttcagaacag	tccgagaaga	agcagacaga	gtgaattttt	ctcgttctgg	2640
tgagggttgg	tctgggtt	ttgttgggtt	cccttctctt	ttaggttact	tcttttacata	2700
tgcagatagg	caacaagact	tttggatgtt	ttattatgtt	tcaagacattt	tcttcaaagc	2760
tgttgcgt	gtgtcgagc	atacccttgc	tggagcttgc	atcatgtatgt	cattttctgt	2820
aggttatttc	catcgatattc	aatgtgagaa	atttccaaac	agtttttacgt	ataagtgtac	2880
tgctgcgtt	agaatattctgg	ctcttttcca	gcaccatgt	ggggtaactt	gaactgttta	2940
ggattatgtt	gttacaaatgg	acggccaccc	gatgtatattt	tattttttttt	acatttcagat	3000
ctttatgtct	aaagcaatcg	aagtcttttt	tgggttccgc	cacgagaaat	aaaaatttga	3060
tcaatcccca	tcatttttgc	aggcagagca	aatgtatgtt	ctcggccatgt	3120	
tcaacagcca	attgtgttcc	gggaaggaaa	ttcgacacaca	gttatactctt	tcaatccatc	3180
agaacagacg	agagaggagg	ttgttgcgtt	ttgtttaac	cgcgctgaaa	tctcggtttt	3240
ggactcaaaac	tggacttgg	tccctagcca	aatttcttctt	gaagtgcagc	atgacgatac	3300
caaacttattc	acggcagac	atcgccttta	ctggaaaagct	tccatcccac	ctcttgggtct	3360
gagaacatata	ttcattgtca	atgggaatgt	cgagtgttgc	aaagctactt	cgtctaaact	3420
caaatacgct	tctgtatgtt	acccttttcc	ttgttccctt	ccatattttt	gctccaaact	3480
ggacaacgac	tttactgttgc	tttggagaaa	atcacccttct	tctcagaaaa	ctcgctttt	3540
cgatcaact	cggaagatag	tccatagaaa	cggtatgttgc	acttgtgttgc	gagaagagat	3600
aggtatgtac	tctgttgc	agagtggggat	ttaccctgtt	aaaccatgtt	gtgaagctca	3660
gccaaattgtt	caacctgtat	gacatgtatgt	caccctgtat	ggtctgttgc	ttcaagaatgt	3720
cttcttttac	cctaaaaacca	aatgggagaa	atcacccttct	tctcagaaaa	ctcgctttt	3780
caactggatgt	aaatacgctt	aggatcaatgt	ggtcgagata	gaatatcatgt	ttgagcttct	3840
tggtaatgtat	tttgcgtt	gggaatttgc	tgtccgttac	aagactgtat	ttgacaacaa	3900
gaagggttttcc	tattcagatc	tcaatgggtt	ccaaatgtatc	aggagagaaaa	tttatgtataa	3960
gatccctttt	caaggaaact	actacccttac	gccatctctc	gtattttatcc	aaggatccaa	4020
tggtcagaga	ttctccgtgc	actctcggtt	tttgcgttgc	tcaaaagagg	4080	

ttgggttggag	attatgtctgg	acagacgggt	ggttcgtat	gacggacggg	gtctaggca	4140
aggtgtatg	gataaccgcg	caatgaccgt	ggtatttcac	cttcttgcgg	aatctaacat	4200
ttctcaagca	gaccctgctt	ccaacactaa	cccgaggaac	ccctcgcttc	tctctcacct	4260
cataggtgt	cactaaact	accccataaa	cacattcatt	gccaagaaaac	cgcaagacat	4320
atctgtcggt	gttcacaaat	acggttcctc	tgctccctta	gccaaaccgt	taceatgtga	4380
cctccacatt	gtaaatttca	aggttcctcg	tccatccaa	tactctcagc	aattggaaaga	4440
agacaaggca	aggttcgctc	ttatcctcaa	tagacgagct	tgggattcag	cttattgcca	4500
taaaggaaaga	caagtaaact	gcacaaggcat	ggctaataa	ccagtaaact	tttccgacat	4560
gttcaaagat	cttgagctt	caaaggtaaa	accaacttca	ctgaatctct	tgcaagaaga	4620
tatggagatt	cttgggtacg	atgaccaaga	gctacctcg	gatagttcac	agccacggga	4680
aggacgtgtc	tcgatctctc	ccatggaaat	acgagcttat	aagcttgaac	tgcgacctca	4740
caagtgaacc	tgctgaagat	ccgctagagt	ccgcaaaaat	caccgtctc	tctctacaaa	4800
tctatctctc	tctattttc	tccagaataa	tgtgtgagta	gttcccagat	aagggattta	4860
gggttcttat	agggttgc	tcatgtgttg	agcatataag	aaacccttag	tatgtatttg	4920
tatgttaaa	atacttctat	caataaaaatt	tctaattccta	aaaccaaaaat	cccgcgagag	4980
acctcttaat	taa					4993

<210> 29
<211> 3825
<212> DNA
<213> hybrid

<400> 29						
ccatggcggag	gatctcgtgt	gacttgagat	ttcttctcat	cccgccagct	ttcatgttca	60
tctacatcca	gatgaggctt	ttccagacgc	aatcacagta	tgcatcg	ctcagttccg	120
ctatcgaatc	tgagaaccat	tgcaatgc	aaatgcgagg	cctcatagat	gaagtttagca	180
tcaaacagtc	gcggattgtt	gccctcgaa	atatgaagaa	ccgcccaggac	gaagaacttg	240
tgcagctaa	ggatctaatac	cagacgttt	aaaaaaaaagg	aatagcaaaa	ctcaactcaag	300
gtggagccat	ggattccaat	tcaggcgc	tcgtttagat	cacaactaaa	gatctatacg	360
ataggatga	gtttcttgc	acagatgg	gtccatggaa	acaagggttg	agagttacgt	420
ataaaagacga	tgagtgggag	aaagagaagc	tcaaaatctt	cgttgc	cattctcata	480
acgatccgg	ttggaaatttgc	actgttaggg	agttatatac	gagacaatcc	agacatattc	540
ttgacacat	tgtttagact	tatctaa	tatgacgaaa	gtttttgtct	ttggttttaa	600
tatTTTta	ctctccat	gttatccct	gaacaatctt	aatgtctta	aaattctcat	660
gacgtcattt	aactctataa	ccaaacttct	ttgtgggtt	ctgtttttt	ttagttcgt	720
gatgaaacag	agttctagaa	gttcgttctt	ttggaaaatt	tgaagtctt	ggagctaaag	780
tttgggggg	tattactggg	tttttagatt	gaaggatagc	tagaatctt	tttgggg	840
ggtttgggg	gaatatgtt	aataggattc	aagaagaaag	tttataatggg	aggagatgtc	900
atatctggag	agatgtgttgc	gagacgttcc	acctaataaa	caagaagctt	tgactaaatt	960
ggtaaaggat	gggcagctag	agattgttg	aggtggctgg	tttatgat	atgaggctaa	1020
ttcacattat	tttgccataa	ttgaacagat	agcagagggt	aatatgtggc	tgaatgacac	1080
aattgggggtt	atttcttctt	attttgggg	ttatagatcc	tttggctatt	catcaaccat	1140
ggcttatctt	ctccggcgta	ttgggttttgc	aaacatgttt	attcaaaagg	ctcattacga	1200
gctcaagaaa	gacccggccc	agcataagaa	tcttgaatat	atttggcg	agagctggga	1260
tgctatggaa	accacagata	tctttgttca	tatgtcg	ttttattcat	acgatatccc	1320
acacacttgt	ggaccagac	ctgcaatttgc	ctgtcagttt	gatttcg	ggatgcgggg	1380
attnaagtat	gaactttgtc	catggggaaa	gcacccatgt	gagaccacac	tagaaaatgt	1440
gcaggagagg	gcattaaagc	ttctggatca	atacaggaaa	aaatccactt	tatatcgac	1500
taatacactt	tttatacctc	ttggatgtt	tttttaggttac	atttagatcg	atgaaggccg	1560
ggctcagttc	cgtaactacc	agatgttttgc	tgatcacatc	aactctaaatc	ctagtctaaa	1620
cgcagaagca	aagtggta	ttttgggg	ttatctcaga	acagtccg	aagaagcaga	1680
cagagtgaat	tattctcg	ctgggttttt	tggcttgcgt	cagggttgc	ttttcccttc	1740
tctgtcagg	gacttctta	catatgcaga	taggaacaa	gactattgg	ttgttatttt	1800
tgtttcaaga	cctttcttca	aagctgttgc	tcgtgtgtc	gagcatacc	ttcgtgg	1860
tgagatcatg	atgtcatttgc	tgcttaggtt	ttgccccatc	attcaatgtt	agaaatttcc	1920
aacaagttt	acgtataat	tgactgtgc	aagaagaaat	ctggcttctt	tccagcacca	1980
tgatggggta	actggacttgc	ctaaggat	tgtgttacaa	gattacggca	cccgatgca	2040
tacttcatttgc	caagacccat	agatctttat	gtctaaagca	atcgaaatgc	ttcttggat	2100
ccgcccacgg	aaagaaaaat	ctgatcaatc	cccatcatttgc	ttcgaggc	agcaaatgg	2160
atcaaagtat	gatgtcg	cagttcaca	gccaatgttgc	ccccggggaa	gaaattcgca	2220
cacagtata	ctcttcaatc	catcagaaca	gacgagagag	gagggtgt	cggttgc	2280
taaccggcg	gaaatctcg	ttttggactc	aaactggact	tgtgtcccttca	gccaatttc	2340
tcctgaatgt	cagcatgac	ataccaaactt	attcaccggc	agacatcgcc	tttactggaa	2400
agcttccatc	ccagcttgc	gtctgagaac	atatttcatttgc	tttactggaa	atgtcg	2460
tgagaaagat	actccgttca	aactcaaaata	cgcttgcag	tttgacccat	ttccttgc	2520
tcctccat	tcctgcttca	aactggacaa	cgacgttact	gagatccgaa	atgaacatca	2580
gactcttgc	tttgatgttgc	agaacggatc	actgcggaa	atagtc	gaaacggatc	2640

agagactgtt	gtgggagaag	agataggtat	gtactctagt	ccagagagtg	gagttacct	2700
gttcaaacc	gatggtaag	ctcagccaat	tgttcaacct	gatggacatg	tagtcaccc	2760
tgagggtctg	ctgggtcaag	aagtcttctc	ttaccctaaa	accaaattgg	agaaatcacc	2820
cctctctcg	aaaactcg	tttacactgg	aggtataatcg	cttcaggatc	aagtggtcg	2880
gatagaat	catgttgc	ttcttggtaa	tgattttgat	gaccggaaat	tgattgtccg	2940
gtacaagact	gatgttgc	acaagaaggt	cttctattca	gatctcaatg	gtttccaaat	3000
gagcaggaga	gaaacttatg	ataagatccc	tcttcaagga	aactactacc	caatgccatc	3060
tctcgattt	atccaaggat	ccaatggta	gagattctcc	gtgcactctc	gtcaatctct	3120
cggtgttgc	agcctcaaag	agggttgggt	ggagattatg	ctggacagac	ggttgggtcg	3180
tgtatgcgga	cgggtctag	ggcaagggtgt	gatggataac	cgcgcaatg	ccgtggatt	3240
tcacccctt	gcggaatcta	acatttctca	agcagaccct	gcttccaaca	ctaaccggag	3300
gaacccttcg	cttcctctc	acctcatagg	tgctactta	aactacccca	taaacacatt	3360
cattgccaag	aaaccgc	acatatctgt	gcgtgttcca	caatacggtt	ccttgcctc	3420
tttagccaaa	ccgttaccat	gtgacccatc	cattgttaat	ttaaagggtt	ctcgccatc	3480
caaatactct	cagaatttgg	aagaagacaa	gccaagggtt	gctcttatcc	tcaatagacg	3540
agcttggat	tcagcttatt	gccataaaagg	aagacaagta	aactgcacaa	gcatggctaa	3600
tgaaccagta	aactttccg	acatgttcaa	agatcttgc	gcttcaaagg	taaaaccaac	3660
ttcactgaat	ctcttgc	aaagatatgga	gattttggg	tacgatgacc	aagagctacc	3720
tcgagatagt	tcacagccac	gggaaggacg	tgtctcgatc	tctcccatgg	aaatacgacg	3780
ttataagctt	gaactgcgac	ctcacaagtg	aacctgctga	agatc		3825

<210> 30
<211> 2181
<212> DNA
<213> hybrid

<400> 30						
ggcgcgcctc	gaggcgatcg	cagatctcat	tataccgtta	gaagcatagt	taaaatctaa	60
agcttgcgt	taattctagt	cattttcat	tgttgggttc	tacattatta	atgaattttc	120
taatgc当地	acagaattt	aatccaaaatt	gttgaattat	gctaaacatg	taacatacgt	180
atatctccgc	cttgcgtgtt	gtttaactt	gaagttatca	taagaaccac	aaataacta	240
gttaatctat	gagaaggcg	gtggcaacac	aaacaagagt	atctaagatt	ttcatttgc	300
actatagaa	tataatatct	cttacatgtat	ttaatgaatc	cacatgttca	cttctcattt	360
gtccacaaga	tcacaacttt	atcttcaata	ttcacaactt	gttataatcc	ccacaatttc	420
attctttca	cttagcccc	caaataactt	tgtccctta	tttgcaccc	tttgcattt	480
atttatttctt	gtggagctaa	gtgttcatat	tattttctt	ctcaaaaaaa	caaaaacaaa	540
aaaaaaagaga	agaaaaccat	ggcgaggatc	tcgtgtgact	ttagatttct	tctcatccg	600
gcagcttca	tgttcatcta	catccagatg	aggctttcc	agacgcatac	acagtatgca	660
gatgcctca	gttccgttat	cgatctcgaa	aaccattgc	ctgtcaat	gcgaggcctc	720
atagatgaag	ttagcatcaa	acagtgcgg	attgttgc	tcaagatata	gaagaaccgc	780
caggacgaag	aacttgc	gcttaaggat	ctaattccaga	cgtttggaaa	aaaaggaata	840
gcaaaactca	ctcaagggtgg	agccatggct	ctaagggtgc	atagaaggaa	ccattttcg	900
cctagaaata	cggatctgtt	cccgatttt	gcaaaagatc	gtgtggttat	cgtcttgc	960
gtgcataatc	gggctcgat	ttttcgatc	acagtgaaa	gtttgtcga	gtttaaagg	1020
ataagtgaga	cattttgtat	gtttagtcat	gatggttact	ttgaagagat	gaataggatt	1080
gtggagagta	ttaatttttg	tcaagtggaa	cagattttct	cgccatttattc	gcctcatata	1140
tatcgacta	gttcccccgg	tgtgaccctg	aatgttgc	agaacaagggg	tgtgaggca	1200
aaggggcatt	gttgcggtaa	tcctgtatcg	tatggaaatc	atcggtctcc	gaagattgt	1260
tcttgcggc	atctctgtt	gtggatgtat	aacactgtat	gggatgggt	ggaagagact	1320
aaaggacatg	agggccat	ccttttcatt	gaaagatc	attttctgtt	tcctaatgccc	1380
tatcgtaaca	tacagactct	tacgaggctg	aaacccgca	agtgtccgt	ctgttttgc	1440
gctaatttag	caccgtctga	tgtgaggatc	agaggagaag	ggcttggaaa	tttgggttgc	1500
gagagaatgg	gaaatgttgg	gtattttttt	aatagaatgt	tgtggggagaa	tattcatcg	1560
aaggcaagag	agttttgttt	cttgcgtat	tacaactggg	atataacat	gtggggcaacg	1620
gttttcccg	cgtttgggtt	cccggtgtac	acattgcgag	ggccttaggac	tagtgcggta	1680
cactttggaa	aatgtgggtt	gcatcaaggt	agaggagatg	agggttattt	catcgataat	1740
ggggtcgtaa	acatagaatg	taaggaaaca	gataaaatgg	tgaacataaa	agaaggatgg	1800
ggagttccgg	tgtataagca	tcaagcggtt	tataaagccg	gtttcgaagg	ttggggaggt	1860
tggggcgat	atagggaccg	acattttatgt	ttggattttgc	ccactatgt	tcgttacagc	1920
atagtcgt	catttccat	aaacggatcc	gctagatgtcc	gcaaaaatca	ccagtcctc	1980
tctacaatc	tatctctctc	tatttttctc	cagaataatg	tgtgaggatgt	tcccagataa	2040
ggaatttagg	gttcttatag	ggtttcgtc	atgtgttgc	catataagaa	accctttagt	2100
tgtattttgt	tttgcataat	acttctatca	ataaaatttc	taatcctaaa	accaaattcc	2160
cgcgagagac	ctcttaattt	a				2181

<210> 31

<211> 1394
<212> DNA
<213> hybrid

<400> 31

ccatggcgag	gatctcgtgt	gacttgagat	ttcttctcat	cccggcagct	ttcatgttca	60
tctacatcca	gatgaggcctt	ttccagacgc	aatcacagta	tgcagatcgc	ctcagttccg	120
ctatcgaatc	tgagaaccat	tgcactagtgc	aatgcgagg	cctcatagat	gaagtttagca	180
tcaaacagtc	gcggattgtt	gccctcgaag	atatgaagaa	ccgcccaggac	gaagaacttg	240
tgcagctaa	ggatctaatac	cagacgtttg	aaaaaaaaagg	aatagcaaaa	ctcactcaag	300
gtggagccat	ggctctaagg	ttgcatagaa	ggaaccattt	ttcgccctaga	aatacgatc	360
tgttcccgga	tttggcaaaa	gatcgtgtgg	ttatcgtctt	gtatgtcat	aatcgggctc	420
agtattttcg	agtacacagt	gaaagtttgt	cgaaggttaa	aggtataagt	gagacattgt	480
tgatttgttgc	tcatgtgtt	tactttgaag	agatgaatag	gattgtggag	agtattaagt	540
tttgtcaagt	gaaaacagatt	ttctcgcctt	attcgcctca	tatataatcg	actagcttcc	600
cgggtgtgac	cctgaatgtat	tgtaagaaca	agggtgtatga	ggcaaagggg	cattgtgaag	660
gtaatcctga	tcagtatggg	aatcatcggt	ctccgaagat	tgtatcttg	aagcatcaact	720
gggtgtggat	gatgaacact	gtatggatg	ggttggaaaga	gactaaagga	catgaggggc	780
atatccttt	cattgaagaa	gatcattttc	tgtttcttaa	tgccatcgat	aacatacaga	840
ctcttacagag	gctgaaaccc	gcaaagtgtc	ctgactgttt	tgctgctaatt	ttagcaccgt	900
ctgatgtgaa	gtcaagagga	gaagggtctg	aaagtttgg	tgcagagaga	atgggaatgt	960
ttgggtatttc	ttttaataga	agtgtgtggg	agaatattca	tcaaggcga	agagagtttt	1020
gtttcttgc	tgattacaac	tgggatataa	cgatgtggc	aacggtttc	ccgtcggttg	1080
gttccccgt	gtacacatttgc	cgagggccta	ggactagtgc	ggtacacttt	ggaaaatgtg	1140
ggttgcatca	aggttagagga	gatgagggtt	attgcatcga	taatggggtc	gtaaacatag	1200
aagttaaagga	aacagataaa	gttgcataaca	taaaagaagg	atggggagtt	cgggtgtata	1260
agcatcaagc	gggttataaa	gccggtttcg	aagttgggg	aggttggggc	gatgataagg	1320
accgacattt	atgtttggat	tttgcacta	tgtatcgat	cagcagtagc	agtgcatctc	1380
	catgaaacgg	atcc				1394

<210> 32
<211> 312
<212> DNA
<213> Arabidopsis thaliana

<400> 32

ccatggcgag	gatctcgtgt	gacttgagat	ttcttctcat	cccggcagct	ttcatgttca	60
tctacatcca	gatgaggcctt	ttccagacgc	aatcacagta	tgcagatcgc	ctcagttccg	120
ctatcgaatc	tgagaaccat	tgcactagtgc	aatgcgagg	cctcatagat	gaagtttagca	180
tcaaacagtc	gcggattgtt	gccctcgaag	atatgaagaa	ccgcccaggac	gaagaacttg	240
tgcagctaa	ggatctaatac	cagacgtttg	aaaaaaaaagg	aatagcaaaa	ctcactcaag	300
gtggagccat	gg					312

<210> 33
<211> 27b
<212> DNA
<213> Glycine max

<400> 33

ccatggcgag	aggaggcaga	tcagtgggta	gcagcagcag	caaatggagg	tactgcaacc	60
cttccttata	cttgcaggcgc	ccaaagcgtc	ttgcgtctgt	tttcatcggt	ttcggttgg	120
tctctttcg	tttctgggac	cgtcaaactc	tcgtcagaga	gcaccagggt	gaaatttctg	180
agctgcagaa	agaagtgtact	gattgaaaa	atttgggtgg	tgatttaaat	aacaaacaag	240
gtggtacctc	tggaaaaact	gacttgggg	ccatgg			27b

<210> 34
<211> 9240
<212> DNA
<213> hybrid

<400> 34

ggcgcgcctc	gaggcgatcg	cagatccgt	ataacaaaaat	ttgaatcgca	cagatcgatc	60
tctttggaga	ttctataacct	agaaaatgg	gacgattttgc	aatatctctgt	aaaaattctg	120
gtttcttctt	gacggaaagaa	gacgacgt	ccaatatttc	ggttagtact	gaaccggaaa	180
gtttgactgg	tgcaccaat	ttaatgtacc	gtacgtaacg	caccaatcgg	attttgtatt	240

caatggccct tatttgtgag cccattaatt gatgtacgg cctaaactaa atccgaacgg 300
tttttttcag cgatccgcga cggtttgtat tcagccaata gcaatcaatt atgttagcagt 360
ggtgatccct gtcaaaccag taaagctaga tctggaccgt tgaattgggt caagaagca 420
catgttgtga tatttttacc cgtacgatta gaaaacttga gaaacacatt gataatcgat 480
aaaaaccgtc cgatcatata aatccgctt accatcggt cctataaatt aatataata 540
gcccgtacag cgtgaagact gacaatatta tcttttgcg attcggagct caagttgaa 600
attcggagaa gcttagagact tttctgatcc ccatggcgag agggacgaga tcagttggta 660
gcagcagcag caaatggagg tactgcaccc ttcgttattt ctgttgcgc ccaaggcgtc 720
ttgctctgt cttcatcggt ttgtttgtt gaaatttcg agctgcagaa agaagtgact gatttggaaa 780
tcgtcagaga gcaccagggtt aacaaacaag gtggtacctc tggggaaact gacttgggaa 840
atttgggtt tgattnaat aacaaacaag gtggtacctc tggggaaact gacttgggaa 900
ccatgggaca gatgcctgtg gctgctgtg tggttatggc ctgcagtcgt gcagactatc 960
ttgaaaggac tgtaaatca gtttaacat atcaaactcc ctgtgttca aaatatcctc 1020
tatttataatc tcaggatggg tctgtatcaag ctgtcaagag caagtcattt agctataatc 1080
aattaacata tatgcagcac ttggattttt aaccagtggt cactgaaagg cctggcgaac 1140
tgactgcgtt ctacaagatt gcacgtcaact acaagtgggc actggaccag ttgttttaca 1200
aacacaaatt tagtcgagtg attatactatc aagatgatata gaaatttgcg ccagacttct 1260
ttgattactt tgaggctgca gctagtcgtt tggataggc taaaaccatt atggctgtt 1320
catcatggaa tgataatggg cagaaggcgtt gggtgatgc tcaagagatc gacttgggat gagttatcac 1380
cagattttt tcctggcctt caaagtggcc aaaggcttac tgggatgatt ggctgagact aaggaaacatacataaaggcc 1440
gccaattcat tcgaccggaa gtctgtagaa catacaattt tggtgaacat gggcttagtt 1500
tgggacagtt ttctagtcag ggttttttttggggaa ggacctgggaa agtctgggat 1560
ggaaagcaaa ggacactgggaa tagtgagaca agcagcagcatacataaaggcc 1620
aggatgtatc tcgtatccgg tttctatccatcataaaggcc 1680
ttggatattt tgaagaatggg ttctatccatcataaaggcc 1740
tggttcgaaat ccgacacaaca ttggatatttgcg tttctatccatcataaaggcc 1800
tttggatattc aaatttctgtt tggatatttgcg tttctatccatcataaaggcc 1860
ctacaatattc atctctctctt tggatatttgcg tttctatccatcataaaggcc 1920
ggaatttaggg ttcttataagg ttggatatttgcg tttctatccatcataaaggcc 1980
gtattnatgtt ttgtaaaata ttggatatttgcg tttctatccatcataaaggcc 2040
gcgcctcgag gcgatcgcaat ttggatatttgcg tttctatccatcataaaggcc 2100
atttttgttt cagtggttac ttggatatttgcg tttctatccatcataaaggcc 2160
aagattattt gttgtatgtt ttggatatttgcg tttctatccatcataaaggcc 2220
ttttaggagat accagccagg ttggatatttgcg tttctatccatcataaaggcc 2280
tcgttatctt ttcatcaaa ttggatatttgcg tttctatccatcataaaggcc 2340
tatgttccgc cgatatgcct ttggatatttgcg tttctatccatcataaaggcc 2400
cgatgttac ccttttccatcataaaggcc 2460
cagaacata taggtattca ttggatatttgcg tttctatccatcataaaggcc 2520
ttccattttt gttatataaa ttggatatttgcg tttctatccatcataaaggcc 2580
tgtacgaat aacttctatt ttggatatttgcg tttctatccatcataaaggcc 2640
gtaccacata taggaaggta ttggatatttgcg tttctatccatcataaaggcc 2700
ccttaccacg aagagataag ttggatatttgcg tttctatccatcataaaggcc 2760
gttaatgata agggattaca ttggatatttgcg tttctatccatcataaaggcc 2820
gccacaggat ccaatggcca ttggatatttgcg tttctatccatcataaaggcc 2880
agatagcaaa caacattata ttggatatttgcg tttctatccatcataaaggcc 2940
catagaagtc catttcctctt ttggatatttgcg tttctatccatcataaaggcc 3000
gtacgacgag cagcaatggg ttggatatttgcg tttctatccatcataaaggcc 3060
gtcttgcct gcttttcattc ttggatatttgcg tttctatccatcataaaggcc 3120
ctctcgtagt agagcaccag ttggatatttgcg tttctatccatcataaaggcc 3180
aaaattttgtt ggatgattta ttggatatttgcg tttctatccatcataaaggcc 3240
ggaccatggg ttccaattca ttggatatttgcg tttctatccatcataaaggcc 3300
ggattgagtt tctgtataca ttggatatttgcg tttctatccatcataaaggcc 3360
aagacgatga gtggggagaaa ttggatatttgcg tttctatccatcataaaggcc 3420
atccctgggtt gaaatttgact ttggatatttgcg tttctatccatcataaaggcc 3480
acaccattgt tgagacttta ttggatatttgcg tttctatccatcataaaggcc 3540
ttttaatttc tcccatgttt ttggatatttgcg tttctatccatcataaaggcc 3600
gtcattaaac tctataacca ttggatatttgcg tttctatccatcataaaggcc 3660
gaaacagagtt tctagaagtt ttggatatttgcg tttctatccatcataaaggcc 3720
gtttttttat tactgggtt ttggatatttgcg tttctatccatcataaaggcc 3780
ttgttttgaa tatgtttaat ttggatatttgcg tttctatccatcataaaggcc 3840
tctggagaga tggggagagtt ttggatatttgcg tttctatccatcataaaggcc 3900
taaggatggg cagctagaga ttggatatttgcg tttctatccatcataaaggcc 3960
acattttttt gccataattt ttggatatttgcg tttctatccatcataaaggcc 4020
tgggttattt cctaagaatt ttggatatttgcg tttctatccatcataaaggcc 4080
ttatctctc cggcgtatgg ttggatatttgcg tttctatccatcataaaggcc 4140
caaqaaqaac cttggccagc ttggatatttgcg tttctatccatcataaaggcc 4200
ttggatatttgcg tttctatccatcataaaggcc 4260
ttggatatttgcg tttctatccatcataaaggcc 4320

tatggaaaacc	acagatatct	ttgttcata	gatgccgtt	tattcatacg	atatccaca	4380
cacttgttga	ccagagcctg	caatttgcgt	tcagttgtat	ttcgctcgga	tgcgggatt	4440
taagtatgaa	ctttgtccat	ggggaaagca	cccagtggag	accacactag	aaaatgtgca	4500
ggagagggca	ttaaagcttc	tggatcaata	caggaaaaaa	tccactctat	atcgaactaa	4560
tacactttt	atacccttgc	gagatgatt	tagtgcatt	agtatcgatg	aagccgaggc	4620
tcagttccgt	aactaccaga	tgttgttga	tcacatcaac	tctaattccta	gtctaaacgc	4680
agaagcaaag	tttggtaactt	tggaggatta	tttcagaaca	gtccgagaag	aagcagacag	4740
agtgaattat	tctcgctctg	gtgagggttg	ctctggtcg	gttgggttgc	tcccttctct	4800
gtcagggtac	ttcttacat	atgcagatag	gcaacaagac	tattggagtg	gttattatgt	4860
ttcaagacct	ttctcaaaag	ctgttgatcg	tgtgctcgag	catacccttc	gtggagctga	4920
gatcatgatg	tcatttctgc	taggttattg	ccatcgaaatt	caatgtgaga	aatttccaac	4980
aagttttacg	tataagtta	ctgctgcaag	aagaatctg	gcttcttcc	agcaccatga	5040
tggggtaact	ggaactgtca	aggatattgt	gttcaagat	tacggcaccc	ggatgcatac	5100
ttcattgcaa	gaccattcaga	tctttatgtc	taaaggaaatc	gaagttcttc	ttgggatccg	5160
ccacgagaaa	gaaaaatctg	atcaatcccc	atcattttc	gaggcagagc	aatgagatc	5220
aaagtatgat	gtccggccag	ttcacaaggc	aattgtgc	cgggaaaggaa	attcgcacac	5280
agttatactc	ttcaatccat	cagaacagac	gagagaggag	gtggtgcgg	ttgtttaa	5340
ccgcgctgaa	atctcggtt	tggactcaa	ctggacttgt	gtccctagcc	aaatttctcc	5400
tgaagtgcag	catgacgata	ccaaactatt	caccggcaga	catcgccctt	actggaaagc	5460
ttccatccca	gctttggtc	tgagaacata	tttcatgtct	aatgggaatg	tcgagtgta	5520
gaaagctact	ccgtttaaac	tcaaatacgc	ttctgagttt	gaccatttc	ttgtcttcc	5580
tccatattcc	tgctccaaac	tggacaacga	cgttactgag	atccggaaatg	aacatcagac	5640
tcttgttgc	gatgtgaaga	acggatcact	gccccggat	gtccatagaa	acggatcaga	5700
gactgttgt	ggagaagaga	taggtatgt	ctcttagtca	gagagtggag	tttacctgtt	5760
caaaccagat	gttgaagctc	agccaaattgt	tcaaccgtat	ggacatgtag	tcacctctga	5820
gggtctgtcg	gttcaagaag	tcttctctta	ccctaaaacc	aatgggaga	aatcaccct	5880
ctctcagaaa	actcgtctt	acactggagg	taatacgctt	caggatcaag	ttgtcgagat	5940
agaatatacat	gttggacttc	ttggtaatga	tttggatgac	cgggaaattga	ttgtccggta	6000
caagactgat	gttgacaaca	agaaggtctt	ctattcgat	ctcaatgggt	tccaaatgag	6060
caggagagaa	acttattgata	agatccctct	tcaaggaaac	tactcccaa	tgccatctct	6120
cgcatttatac	caaggatcca	atggtcagag	atttccgtg	cactctcgat	aatctctcg	6180
tgttgcgaagc	cttccaaagagg	gttgggttgc	gattatgtc	gacagacgg	ttgttgcgt	6240
tgacggacgg	ggtctaggc	aagggtgtat	ggataaccgc	gcaatgaccg	ttgttattca	6300
ccttcttgcg	gaatctaaca	tcttctcaagc	agaccctgct	tccaaacacta	acccgaggaa	6360
cccttcgcctt	ctctctcacc	tcataggtc	tcacttaaac	tacccataaa	acacattcat	6420
tgccaagaaa	ccgcaagaca	tatctgtgc	tggttccacaa	tacggttcc	ttgctccctt	6480
agccaaaccg	ttaccatgt	acctccacat	tgtaaatttc	aagggttcc	gtccatccaa	6540
atactctcag	caatttggaa	aagacaagcc	aagggtcgct	tttacccctca	atagacgagc	6600
ttgggatca	gcttattgtcc	ataaaaggaa	acaagtaaac	tgcacaagca	tggctaatga	6660
accagtaaac	ttttccgaca	tgttcaaaa	tcttgcagat	tcaaaggtaa	aaccaacttc	6720
actgaatctc	tttgcagaag	atatggagat	tcttgggtac	gatgaccaag	agcttacctg	6780
agatagttca	caggccacgg	aaggacgtgt	ctcgatctt	cccatggaaa	tacgagctt	6840
taagtttgc	ctgcgaccc	acaagtgaac	ctgtcaaga	tccgcttagag	tccgcaaaaa	6900
tcaccagatct	ctctctacaa	atctatctt	ctcttatttt	ctccagaata	atgtgtgagt	6960
agttcccaga	taagggaaatt	agggttctta	taggtttcg	ctcatgtgtt	gagcatataa	7020
gaaaccctta	gtatgttatt	gtatttgtaa	aataacttcta	tcaataaaaat	ttctaattct	7080
aaaaccaaaa	tcccgccg	gcctcgaggc	gatcgagat	ctcattatac	cgtagaaagc	7140
atagttaaaa	tctaaagctt	gtcgttaatt	ctagtcatt	tacattgtt	ggttctacat	7200
tattaatgaa	ttttctaatg	caaatacaga	atttaaatca	aaattgttga	attatgtcaa	7260
acatgttaaca	tactgtatc	tccggcttgc	gttgggttgc	taacttgcatt	tatcataaga	7320
accacaaaata	cacttagaa	tctatgagaa	ggcagggtggc	aacacaaaaca	agagtatcta	7380
agattttcat	tttgcgtat	aggaatataa	tatctttat	ctgatattaa	gaatccacat	7440
gttcacttct	catttgtcc	caagatcaca	actttatctt	caatattcac	aacttgcatt	7500
atccaccaca	atttcattct	tttcaacttag	ccccacaaaa	tactttgtcc	ccttatttgc	7560
cacccccc	attnaattta	ttcttgc	gctaagtgtt	catattattc	ttcttctcaa	7620
aaaaaccaaaa	acaaaaaaaaa	agagaagaaa	accatggcga	gaggggagcag	atcagtgggt	7680
agcagcagca	gcaaatggag	gtactgcaac	ccttccatt	acttgcggc	cccaaaagcgt	7740
cttgcgtctgc	tcttcattcg	tttgcgttgc	tttttttgcg	tttttttgcg	ccgtcaaaact	7800
ctcgtcagag	agcaccaggt	tggatgttgc	gagctgcaga	aagaagtgc	tgatttggaa	7860
aatttgggttgc	atgatattaa	taacaaacaa	ggtggatct	ctggggaaaac	tgacttgggg	7920
accatggctc	taaggttgc	tagaaggaac	cattttcgc	ctagaaatac	ggatctgttc	7980
ccggatttgg	caaaagatcg	tgtgggtatc	gtctgtatg	tgcataatcg	ggctcagtat	8040
tttcgagtca	cagtgaaag	tttgcgtat	gtttaaggta	taagtgcgg	attgttgatt	8100
gttagtcatg	atgatgtactt	tgaagagatg	aataggattg	tggagagtgat	taagttttgt	8160
caagtggaaac	agattttctc	gccttattcg	cctcatat	atcgatctat	tttccgggt	8220
gtgaccctgt	atgatgttgc	gaacaagggt	gatggggc	agggggatttgc	tgaaggtaat	8280
cctgtatgt	atgatgtatca	tcgggtctcc	aaggatgtat	tttgcgtat	tcacttgcgt	8340
tggatgtat	acactgtat	ggatgggttgc	gaagagacta	aaggacatga	ggggcatatc	8400

cttttcattg	aagaagatca	ttttctgtt	cctaattgcct	atcgtaacat	acagactctt	8460
acgaggcgtga	aaccggcaaa	gtgtcctgac	tgttttgctg	ctaatttagc	accgtctgat	8520
gtgaagtcaa	gaggagaagg	gcttggaaagt	ttgggtgcag	agagaatggg	aatgttggg	8580
tattcttta	ataagaatgt	gtgggagaat	attcatcaga	aggcaagaga	gttttggttc	8640
tttgatgtt	acaactggga	tataacgtat	tggggacgg	tttcccgtc	gtttgggttc	8700
ccggtgtaca	cattgcgagg	gccttaggact	agtgcggat	actttggaaa	atgtgggttg	8760
catcaaggt	gaggagatga	gggtgattgc	atcgataatg	gggtcgtaaa	catagaagtt	8820
aaggaaacag	ataaaagtgt	gaacataaaa	gaaggatggg	gagttcggtt	gtataaagcat	8880
caagcgggtt	ataaaagccgg	tttcgaaggt	tggggaggtt	ggggcgatga	tagggaccga	8940
catttatgtt	tggattttgc	cactatgtat	cgttacagca	gttagcagtgc	atctccatga	9000
aacggatccg	ctagagtccg	caaaaatcac	cagtctctct	ctacaaatct	atctctctct	9060
attttctcc	agaataatgt	gtgagtagtt	cccagataag	ggaatttaggg	ttcttatagg	9120
gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	gtattttgtat	ttgtaaaata	9180
cttctatcaa	taaaaatttct	aatcctaaaa	ccaaaatccc	gcgagagacc	tcttaattaa	9240

<210> 35
<211> 2180
<212> DNA
<213> hybrid

<400> 35	60					
ggcgcgcctc	gaggcgatcg	cagatccat	ataacaaaat	ttgaatcgca	cagatcgatc	120
tctttggaga	ttctataacct	agaaaatgg	gacgatttc	aaatctctgt	aaaattctg	180
gtttcttctt	gacggaaaga	gacgacgact	ccaatattc	ggttagtact	gaaccggaaa	240
gtttgactgg	tgcaaccaat	ttaatgtacc	gtacgtaacg	caccaatcg	attttgtatt	300
caatgggcct	tatctgttag	cccattaatt	gatgtacgg	cctaaactaa	atccgaacgg	360
tttatttcag	cgatccgcga	cgtttgtat	tcagccaata	gcaatcaatt	atgtacgt	420
ggtgatcctc	gtcaaaccag	taaagctaga	tctggaccgt	tgaattgggt	caagaaagca	480
catgttgta	tatttttacc	cgtacgat	gaaaacttga	gaaacacatt	gataatcgat	540
aaaaaccgtc	cgatcatata	aatccgcctt	accatcggt	cctataaaatt	aatatcaata	600
gccgtacacg	cgtgaagact	gacaatatt	tcttttgcg	attcggagat	caagttgaa	660
attcggagaa	gctagagat	tttctgat	ccatggcgag	ggggagcaga	tcagtggta	720
gcacgacg	caaatggagg	tactgcAAC	cttccttata	cttgaagcgc	ccaaagcgtc	780
ttgctctgt	tttcatcggt	ttcggttgcg	tctcttgcgt	tttctgggac	cgtcaactc	840
tcgtcagaga	gcaccagggt	gaaatttctg	agctgcagaa	agaagtgact	gatttgaaaa	900
atttgggtga	tgattnaat	aacaaacaag	gtggtacctc	tggggaaaact	gacttgggga	960
ccatgggaca	gatgcctgt	gctgctgt	tggttatggc	ctgcagtcgt	gcagactatc	1020
ttgaaaggac	tgttnaatca	gtttaacat	atcaaactcc	cgttgcctca	aatatccctc	1080
tatttatatc	tcaggatgg	tctgtatca	ctgtcaagag	caagtcat	agctataatc	1140
aattaacata	tatgcagcac	ttggatttt	aaccagtgtt	cactgaaagg	cctggcgaac	1200
tgactgcgt	ctcaagat	gcacgtact	acaagtgggc	actggaccag	ttgttttaca	1260
aacacaaatt	tagtcgtat	attatactag	aagatgtat	ggaaattgtc	ccagacttct	1320
ttgattactt	tgaggctgc	gctagtctca	tggataggaa	taaaaccatt	atggctgttt	1380
catcatgaa	tgataatgg	cagaaggact	ttgtgcata	tccctatcg	ctataccgt	1440
cagattttt	tcctggcctt	gggtggatgc	tcaagagatc	gacttggat	gagttatcac	1500
caaagtggcc	aaaggcttac	tgggatgatt	ggctgagact	aaaggaaaac	cataaaggcc	1560
gccaattcat	tcgaccggaa	gtctgtagaa	catacaattt	tggtgaacat	gggtctagtt	1620
tgggacagtt	tttcaagtat	tatctggaa	ctataaagct	aaacgtatgt	acgggttgact	1680
ggaaagcata	ggacctggga	tacctgacag	agggaaacta	taccaagatc	ttttctggct	1740
tagtgagaca	agcagcagaca	attcaaggat	ctgaccttgc	ctttaaggct	caaaacataa	1800
aggatgtat	tcgtatccg	tataaaagacc	aagtagatgt	tgaacgcatt	gcaggggaat	1860
ttggatataat	tgaaaatgg	aaggatgggt	tgcctcgaa	agcatataaa	ggagtagtgg	1920
tgtttcgat	ccagacaaca	agacgttat	tcctgggtgg	gccagattct	gtatgcagc	1980
ttggaaatcg	aaatccat	tgcggatcc	ctagagtccg	ccaaaatcac	cagtctct	2040
ctacaaatct	atctctct	attttctcc	agaataatgt	gtgagtagtt	cccagataag	2100
ggaatttaggg	ttcttatagg	gttgcgtca	tgtgttgagc	atataagaaa	cccttagtat	2160
gtattttgtat	ttgtaaaata	tttctatcaa	taaaatttct	aatcctaaaa	ccaaaatccc	2180

<210> 36
<211> 615
<212> DNA
<213> *Arabidopsis thaliana*

<400> 36	60					
ggatccgata	taacaaaatt	tgaatcgac	agatcgatct	ctttggagat	tctatactca	60

gaaaatggag acgattttca aatctctgt aaaaattctgg tttcttcttg acggaagaag	120
acgacgactc caatatttcg gtttagtactg aaccggaaag tttgactgg gcaaccatt	180
taatgtaccg tacgtaacgc accaatcgga ttttgtattc aatgggcctt atctgtgagc	240
ccattaatg atgtgacggc ctaaactaa tccgaacggg ttatccagc gatccgcac	300
ggtttgatt cagccaaatag caatcaatta tggtagcagt gtgatccctg tcaaaccagg	360
aaagcttagat ctggaccgtt gaattgggtc aagaagcac atgttgtgat attttaccc	420
gtacgattag aaaactttag aAACACATTG ataatcgata AAAACCGTCC gatcatataa	480
atccgctta ccacgttgc ctataaattt atatcaatag ccgtacacgc gtgaagactg	540
acaatattat cttttcgaa ttccggagctc aagttgaaa ttccggagaa cttagagat	600
ttctgataac catgg	615

<210> 37
<211> 1319
<212> DNA
<213> hybrid

<400> 37	
ccatggcgag agggagcaga tcagtggta gcagcagcag cAAATGGAGG tactgcaacc	60
cttccttata cttgaagcgc ccaaagcgtc ttgctctgt cttcatcggtt ttcgtttgt	120
tctctttcgt tttctggac cgtcaaaactc tcgtcagaga gcaccagggtt gaaatttctg	180
agctgcagaa agaagtact gattgaaaa atttgggtt gatTTAAAT aACAAACAAG	240
gtggtaccc tggggaaaact gacttgggaa ccatgggaca gatgcctgtt gctgctgt	300
tgggtatggc ctgcagtctgtt gcaagactatc ttgaaaggac tgTTAAATCA gttttaacat	360
atcaaactcc ctgttgcctca aaatatccctc tattttatc tcaggatggg tctgatcaag	420
ctgtcaagag caagtattt gactataatc aattaacata tatgcagcac ttggatttt	480
aaccagggtt cactgaaagg cctggcgaac tgactgcgtt ctacaagatt gcacgtcact	540
acaagtgggc actggaccag ttgttttaca aacacaaatt tagtgcagtg attatactag	600
aagatgatat ggaaatttgcg ccagacttctt ttgattactt tgaggctgc gctagtc	660
tggatagggg taaaaccatt atggctgtt catcatggaa tgataatggg cagaaggcgt	720
ttgtgcatgt tccctatgcg ctataccgtat cagattttt ccctggccctt gggggatgc	780
tcaagagatc gacttgggat gagttatcac cAAAGTGGGC AAAGGCTTAC tgggtatgtt	840
ggctgagact aaaggaaaac cataaaaggcc gccaattcat tcgaccggaa gtctgt	900
catacaattt tggtaacat gggcttagtt tgggacagtt tttcagtcgtatctggaaac	960
ctataaaatc aaacgtatgtg acgggttactt gggaaagcaaa ggacctggg tacctgacag	1020
agggaaaacta taccatgtt ttttctggct tagtgcagaca agcacgcacca attcaaggtt	1080
ctgaccttgc cttaaaggct caaaacataa aggatgtatc tgcgtatccgg tataaagacc	1140
aagtagagtt tgaacgcatt gcaggggaaat ttggatatt tgaagaatgg aaggatgggt	1200
tgcctcgaaac agcatataaa ggagtagtgg tgTTTCGAAT ccagacaaca agacgtgtat	1260
tcctgggttgg gcccagattct gtaatgcgc ttggaaattcg aaatttctgt tgcggatcc	1319

<210> 38
<211> 4957
<212> DNA
<213> hybrid

<400> 38	
ggcgcgcctc gaggcgatcg cagatctaattt ctaaccaattt acgatacgct ttgggtacac	60
ttgatTTTG tttcgtgtt tacatataatc ttgtttata ttgttatctt aaggatctgc	120
acaaagattt tttttgtatg ttcttgatgg ggctcagaag atttgcattt atacactcta	180
atctttagga gataccagcc aggattatatt tcagaagac aatcaaattt tacgtttca	240
aactcggtat cttttcatcc aaaggatgg ccagaatctt tatagaatgtt ttgcattca	300
gaatatgttc ggcgatatg cctttgttgg cttcaatattt ctacatataca cacaagaatc	360
gaccgtattt taccctcttt ccataaaagga aaacacaata tgcagatgtt tttttccac	420
atgcgatgtt atatacgat tcaaaaaatgg ctaaaaagaag ttggataaca aattgacaac	480
tattttcatt tctgttatattt aaatttcaca acacacaaaa gcccgttacatc aagagtctgc	540
ccatgtacgtt aataacttctt atttttggt attggccctt agcccagctc agagtacgt	600
ggggtaaccat atataggaaat gtaacaaat tgcacatggt agccccatataa cgtaccagcc	660
tctccttacc acgaaagatg agatataatg accccacctg ccacgtgtca catcgat	720
gtggtaatgtt ataaaggattt acatccattt atgtttgtgg acatgtatc tggataatgt	780
ttagccacat gatccatgg ccacaggaac gtaagaatgtt agatagattt gattttgtcc	840
gttagatgtt aacaaacattt ataaaagtg tgcgtatcaata ggaactaattt cactcattgg	900
attcatagaa gtccatttcctt cctaagtatc tagaaaccat ggcgagaggg agcagatcg	960
tgggtacccat cagcagcaaa tggaggtactt gcaacccttc ctattactt aagcggccaa	1020
agcgtttgc tctgtcttgc atcggtttcg tttgtgtctc ttccgttttc tgggaccgtc	1080
aaactctcgat cagagagcac cagggttggaaat ttccgtatgtt gcaagaaagaa gtgcgtt	1140
tggatgttggat ttaataaca aacaagggtgg tacctctggg aaaactgt	1200

tggggaccat	ggattccaat	tcagggcgccg	tcgttgatat	cacaactaaa	gatctatacg	1260
ataggattga	gtttcttgc	acagatgggt	gtccatggaa	acaagggttg	agagttacgt	1320
ataaagacga	tgagttggag	aaagagaagc	tcaaaatctt	cgttgttcct	cattctcata	1380
acgatccgg	ttggaaattt	actgttaggg	agtattatca	gagacaatcc	agacatattc	1440
ttgacaccat	ttgtgagact	tttatctaagg	tatgacggaa	gtttttgtt	ttggttttaa	1500
tatTTTaaat	ctctccatg	gttatcccgt	gaacaatctt	aaatgtctta	aaattctcat	1560
gacgtcatta	aactctataa	ccaaacttctt	ttgctgggtt	ctgtttttt	ttagtttgcgt	1620
gatgaaacag	agttctagaa	gttcgttctt	ttggaaaatt	tgaagtcttt	ggagctaaag	1680
tttggttttt	tattactggg	tttgagattt	gaaggatagc	tagaatctta	tttgcgttggg	1740
ggtttggttt	gaatatgtt	aataggattt	aagaagaaaag	tttatatggg	aggagatgtc	1800
atatctggag	agatgggtga	gagacgcttc	acctaataaa	caagaagctt	tgactaaatt	1860
ggtaaggat	gggcagctag	agattgttgg	aggtggctgg	gttatgaatg	atgaggctaa	1920
ttcacattat	tttgcattaa	ttgaacagat	agcagagggt	aatatgtggc	tgaatgacac	1980
aattgggggtt	attcattaaaga	attcttgggg	tatagatccc	tttgcattt	catcaaccat	2040
ggcttatctt	ctccggcgta	ttgggttttga	aaacatgctt	attcaaagga	ctcattacga	2100
gctcaagaaaa	gaccttgc	cc	agcataagaa	tcttgaatat	atttgcgtc	2160
tgctatggaa	accacagata	tctttgttca	tatgatgccg	tttattatcat	acgatatccc	2220
acacacttgt	ggaccagagc	ctgcaattt	ctgtcagttt	gatttcgctc	ggatgcgggg	2280
attnaagttt	gaactttgtc	catggggaaa	gcacccagt	gagaccacac	tagaaaatgt	2340
gcaggagagg	gcattaaagc	ttctggatca	atacaggaaa	aatccactc	tatatcgaac	2400
taatacactt	tttatacctc	ttggagatga	ttttaggtac	attgtatcg	atgaagccga	2460
ggctcagttc	cgttaactacc	agatgttgg	tgatcacatc	aactctaattc	ctagtctaaa	2520
cgcagaagca	aagtttggta	ctttggagaa	gcacccagt	gagaccacac	aagaagcaga	2580
cagagtat	tattctcg	ctgggtggat	ttggcttgcgt	cagggttgg	gtttcccttc	2640
tctgtcagg	gacttcttta	catatgcaga	taggaacaa	gactattgg	gtgggttatta	2700
tgttcaaga	ccttcttca	aagcttggta	tcgtgtgctc	gagcataacc	ttcgtggagc	2760
tgagatcatg	atgtcattt	tgcttaggtt	ttgccatcga	attcaatgtt	agaaaatttcc	2820
aacaagttt	acgtataagt	tgactgtc	aagaagaaat	ctggcttctt	tccagcacca	2880
tgatgggta	acttggact	ctaaggat	tgtggtacaa	gattacggca	cccgatgca	2940
tacttcattt	caagaccttc	agatctttat	gtcttaagca	atcgaagttt	ttcttggat	3000
ccgccccac	aaagaaaaat	ctgtcaatc	cccatcattt	ttcgaggcag	agcaaatttgc	3060
atcaaagttt	gtatgtcg	cagtccacaa	gccaatttgc	gcccggggaa	gaaatttgc	3120
cacgttata	ctcttcattt	catcagaaca	gacgagagag	gagggttgg	cggttgg	3180
taaccgcgt	gaaatctcg	tttggactc	aaacttggact	tgtgtccctt	gccaatttgc	3240
tccttgaatg	cagcatgacg	ataccaaact	atttaccggc	agacatcgcc	tttacttggaa	3300
agcttccatc	ccagcttgc	gtctgagaac	atatttattt	gctaatttgg	atgtcgatg	3360
tgagaaagct	actccgtcta	aactcaata	cgcttctg	tttgacccat	ttcccttgc	3420
tcctccat	tcctgttca	aacttggacaa	cgacgttact	gagatccgaa	atgaacatca	3480
gactcttgc	tttggatgt	agaacggatc	acttgcggaa	atagtccata	gaaacggatc	3540
agagactgtt	gtggggagaag	agataggtat	gtactctgt	ccagagatg	gagtttac	3600
gttcaaaatc	gtatgtcg	tttggatgt	tttgcattt	gatggatgt	tagtccatc	3660
tgagggtctg	ctgggttca	aaatgttctc	tttgcattt	accatgggg	agaaatccacc	3720
cctctctcg	aaaatctcg	tttgcattt	tttgcattt	tttgcattt	tttgcattt	3780
gatagaat	catgttgc	tttgcattt	tttgcattt	tttgcattt	tttgcattt	3840
gtacaagact	gtatgtgaca	acaagaaggt	tttgcattt	gtatctcaat	tttgcattt	3900
gagcaggaga	gaaactttag	ataagatccc	tcttcaagga	aactactacc	caatgc	3960
tctcgat	atccaaggat	ccaatggtca	gagattctt	gtgcactt	gtcaatctt	4020
cggtgttgc	agccctaaag	agggttgg	ggagattat	ctggacagac	gggtgg	4080
tgatgtac	cggggtctcg	ggcaagggt	gtatgtat	ccgcgaat	ccgtgg	4140
tcacccctt	gcggatatt	acatttctt	acggaccc	gcttccaa	ctaacc	4200
gaaccctcg	cttctcttc	acctcat	tgcttactt	aacttccca	taaacacat	4260
cattgccaag	aaaccgc	acatatctt	tgcttactt	caatacg	ttttgttcc	4320
tttagccaa	ccgttaccat	gtgaccttca	cattgtaaat	ttcaagg	tcgtccatc	4380
caaatactt	cagaatttgg	aagaagacaa	gccaagg	gcttccat	tcaatagac	4440
agcttggat	tcagcttatt	gccat	aaagg	aaagacaat	gtatggctaa	4500
tgaaccat	aactttcc	acatgttca	agatcttgc	gttcaaa	taaaaccaac	4560
ttcaactgat	cttcttgc	aagatatgg	gatttgc	tgatgt	aagatgtt	4620
tcgagatgt	tcacagcc	gggaaggac	tgttgc	tctccat	aaatagc	4680
ttataagtt	gaacttgc	ctcacaat	aaatgt	tttgc	tttgc	4740
aaatcacc	tctcttca	caatctt	tctctt	tttgc	tttgc	4800
agatgttccc	agataagg	atttgg	ttatagg	tttgc	tttgc	4860
taagaaaccc	tttagtat	tttgc	tttgc	tttgc	tttgc	4920
cctaaaaacca	aaatcc	ccgc	agagac	tttgc	tttgc	4957

<210> 39
<211> 921
<212> DNA

<213> Chrysanthemum x morifolium

<400> 39

agatctaattc	taaccaatta	cgatacgctt	tgggtacact	tgattttgt	ttcagtggtt	60
acatatatct	tgtttatat	gctatctta	aggatctgc	caaagattat	ttgttgatgt	120
tcttgatgg	gctcagaaga	tttgatatga	tacactctaa	tctttaggag	ataccagcca	180
ggattatatt	cagtaagaca	atcaaatttt	acgtgtcaa	actcgttatc	tttcattca	240
aaggatgagc	cagaatctt	atagaatgat	tgcaatcgag	aatatgttcg	gccgatatgc	300
ctttgttgc	ttcaatattc	tacatatcac	acaagaatcg	accgtattgt	acccttttc	360
cataaaggaa	aacacaatat	gcagatgctt	ttttccaca	tgcaatcaca	tataggtatt	420
caaaaatggc	taaaagaagt	ttggataacaa	attgacaact	atttccattt	ctgttatata	480
aatttcacaa	cacacaaaag	cccgtaatca	agagtctgc	catgtacgaa	ataacttcta	540
ttatggta	ttgggcctaa	gcccgagctca	gagtacggtg	gggtaccaca	tataggaagg	600
taacaaaata	ctgcaagata	gccccataac	gtaccgcct	ctcccttacca	cgaagagata	660
agatataaga	cccaccctgc	cacgtgtcac	atcgcatgg	tgtttaatga	taagggatta	720
catccttcta	ttttgtgga	catgtgcat	gtaatgtcat	gagccacagg	atccaatggc	780
cacaggaacg	taagaatgta	gatagattt	attttgtccg	ttagatagca	aacaacatta	840
taaaaggtgt	gtataatag	gaaactaattc	actcattgga	ttcatagaag	tccattcctc	900
ctaagtatct	agaaaccatg	g				921

<210> 40

<211> 3789

<212> DNA

<213> hybrid

<400> 40

ccatggcgag	agggaggcaga	tcaagtggta	gcagcagcag	caaatggagg	tactgcacc	60
cttccttata	cttgaagcgc	ccaaagcgtc	ttgctctgc	cttcatcggt	ttcggttgc	120
tctctttcgt	tttctggac	cgtcaaaactc	tcgtcagaga	gcaccagggt	gaaatttctg	180
agctgcagaa	agaagtact	gatttggaaa	atttgggtga	tgattttaaat	aacaaaacaag	240
gtggtaccc	tggggaaact	gacttgggg	ccatgattc	caattcaggc	gccgtcggt	300
atatacacaac	taaaatcttca	taatgtatgg	ttgaggatct	tgtatcagat	ggtggtccat	360
ggaaacaagg	ttggagagtt	acgtataaag	acgatgatgt	ggagaaaaag	aagctaaaa	420
tcttcgtgt	tcctcattct	cataacgatc	ctgggtggaa	attgactgt	gaggagatt	480
atcagagaca	atccagacat	attcttgaca	ccatttggta	gactttatct	aaggtatgac	540
gaaaagtttt	gcttttgggtt	ttaatatttt	aattctctcc	catggttatc	ccgtgaacaa	600
tcttaaatgt	cttaaaattt	tcatgacgtc	attaaactct	ataaaccacac	ttctttgctg	660
ggttctgtt	tttttagtt	tcgtgatgaa	acagagttct	agaagttcg	tctttttggaa	720
aatttgaagt	ctttggagct	aaagttttt	tttttattac	ttgggtttga	atttgaagga	780
tagctagaat	ttttttgtt	ttgggggtt	ttttgatatt	ttttaatagg	attcaagaag	840
aaagttata	ttggggggaga	tgtcatatct	ggagagatgg	ttggagagacg	tttcacctaa	900
taaacaagaa	gcttgcata	atttggtaa	ggatgggcag	ctagagattt	ttggagggtgg	960
ctgggttat	aatgtgagg	ctaattcaca	tttatttgc	ataattgac	agatagcaga	1020
gggtatatg	tggctgaatg	acacaatttg	ggtttattct	agaatttctt	gggctataga	1080
tccctttgc	tatttcatcaa	ccatggctt	tcttccgg	cgtatgggt	ttgaaaacat	1140
gcttatttca	aggacttatt	acgagctcaa	gaaagacctt	gcccgatata	agaatcttga	1200
atataatttg	cgtcagagct	ggggatgtat	ggaaaccacac	gatatttttgc	tttcatatgt	1260
gcccgtttat	tcatacgtat	tcccacacac	ttgtggacca	gaggcctgc	tttgcgtca	1320
ttttgtat	gctcgatgc	ggggatttt	gtatgactt	tttccatggg	gaaagcacc	1380
agtggagacc	acatagaaaa	atgtcgagaa	gagggcattt	aagttctgg	atcaatacag	1440
aaaaaaatcc	actttttatc	gaactaatac	acttcttata	ccttttggag	atgatttttg	1500
gtacattatgt	atcgatgaa	ccgaggctca	gttccgttac	taccagatgt	tttttgcata	1560
catcaactct	aatccatgtc	taaacgcaga	agcaaaatgtt	ggtactttgg	aggattattt	1620
cagaacagtc	cgagaagaag	cagacagat	gaattattct	cgtcctgggt	aggttggctc	1680
tggtcaggtt	tttggtttcc	tttctctgc	agggtacttc	tttacatatg	cagataggca	1740
acaagactat	tggatgggtt	attatgtttc	aaagcccttc	tccaaagctg	tttgcgtgt	1800
gtctcgagat	acccttcgtt	gagctgagat	catgtatgtca	tttctgtatg	tttatttgc	1860
tcgaatttca	tgtgagaaat	tttcaacaag	ttttatgtat	tttttgcata	tttgcgtat	1920
aaatctggct	cttttccagc	accatgtatgg	ggtaacttgg	actgtcttgc	tttgcgtat	1980
acaagattac	ggccacccgg	tgcatacttc	atttgcacat	tttgcgtat	tttgcgtat	2040
agcaatcgaa	gttcttcttgc	ggatccgc	cgagaaaagaa	aaatctgtatc	aatccccatc	2100
atttttcgat	gcagagcaaa	tgagatcaaa	gtatgtatgt	tttgcgtat	tttgcgtat	2160
tgctgcccgg	gaaggaaatt	cgcacacagt	tataacttgc	aatccatcg	aacagacgag	2220
agaggaggtt	gtgacgggtt	ttttaaccc	cgctgaaatc	tttgcgtat	tttgcgtat	2280
gacttgtgtc	cctagccaaa	tttctctgtca	agtgatgtat	tttgcgtat	tttgcgtat	2340
cgccagacat	cgcccttact	ggaaagcttc	catcccgat	tttgcgtat	tttgcgtat	2400
cattgctaat	gggaatgtcg	agtgtgagaa	agctactccg	tctaaactca	aatacgctc	2460

ttaggtttgac	ccatttcctt	gtcctcctcc	atattctgc	tccaaactgg	acaacgacgt	2520
tactgagatc	cggaaatgaac	atcagactct	tgtgtttgat	gtgaagaacg	gatcactgcg	2580
gaagatagt	cataaaaaacg	gatcagagac	tgttgtggg	gaagagatag	gtatgtactc	2640
tagtccagag	agtggagctt	acctgttcaa	accagatgtt	gaagctcagc	caattgtca	2700
acctgtatgg	catgtatca	cctctgaggg	tctgtgggt	caagaagtct	tctttaccc	2760
taaaaacaaa	tgggaaaaat	cacccctctc	tcagaaaact	cgtctttaca	ctggaggtaa	2820
tacgcttcag	gatcaagtgg	tcgagataga	atatcatgtt	gagcttctt	gtaatgattt	2880
tgatgaccgg	gaattgattt	tccggtacaa	gactgtatgtt	gacaacaaga	aggcttcta	2940
ttcagatctc	aatggttcc	aatgagcag	gagagaaaact	tatgataaga	tccctttca	3000
aggaaaactac	tacccaatgc	catctctcg	atttatccaa	ggatccaatg	gtcagagatt	3060
ctccgtgcac	tctcgtaat	ctctcggtgt	tgcaagcctc	aaagagggtt	ggttggagat	3120
tatgctggac	agacgggtgg	tccgtatgt	cgaggggggt	ctagggcaag	gtgtgatgga	3180
taaccgcgc	atgaccgtgg	tatttcacct	tcttgcggaa	ttaacattt	ctcaagcaga	3240
ccctgcttcc	aacactaacc	cgaggaaccc	tctgccttc	tctcacctca	taggtgtca	3300
cttaaactac	cccataaaca	cattcattgc	caagaaaccg	caagacatat	ctgtgcgtgt	3360
tccacaatac	ggttcctttt	tccttttagc	caaaccgtt	ccatgtgacc	tccacattgt	3420
aaatttcaag	gttccctgtc	catccaaata	ctctcagcaa	ttggaaagaag	acaagccaag	3480
gttcgcctt	atcccaata	gacgagctt	ggattcagct	tattgccata	aaggaagaca	3540
agtaaactgc	acaagcatgg	ctaatacacc	agtaaacttt	tccgacatgt	tcaaagatct	3600
tgcagcttca	aaggtaaaac	caacttact	gaatctttt	caagaagata	tggagattct	3660
tgggtacat	gaccaagagc	tacctcgaga	tagttcacag	ccacgggaag	gacgtgtctc	3720
gatctctccc	atggaaatac	gagttataa	gcttgaactg	cgacctcaca	agtgaacctg	3780
ctgaagatc						3789

<210> 41
<211> 2145
<212> DNA
<213> hybrid

<400> 41						
ggcgcgcctc	gaggcgatcg	cagatctat	tataccgtta	gaagcatagt	taaaatctaa	60
agcttgtcgt	taattctagt	cattttacat	tgttgtggttc	tacatttatta	atgaattttc	120
taatgc当地	acagaatttta	aatccaaaatt	gttgaattat	gctaaacatg	taacatacgt	180
atatctccgc	cttgtgtgtt	gtattaaactt	gaagttatca	taagaaccac	aaatacacta	240
gtaaaatctat	gagaaggcag	gtggcaacac	aaacaagagt	atctaagatt	ttcattttgt	300
actatagaa	tataatatct	cttatctgt	ttaatgaatc	cacatgttca	tttctcattt	360
gtccacaaga	tcacaacttt	atcttcaata	ttcaccaactt	gttatatatcca	ccacaatttc	420
attcttttca	cttagccccca	caaaataactt	tgtcccttta	tttgcaccc	tttgcaccc	480
atttattttt	gtggagctaa	gtgttcatat	tattttctt	ctcaaaaaaaaa	aaaaacaaaa	540
aaaaaagaga	agaaaaaccat	ggcggagggg	agcagatcag	ttggtagcag	cagcagaaaa	600
tggaggtact	gcaacccttc	ctattactt	aagcgccaa	agcgttgc	tctgctttc	660
atcggtttcg	tttgtgtctc	tttcgtttt	ttggaccgtc	aaactctcg	cagagagcac	720
cagggtgaaa	tttctgagct	gcagaaagaa	gtgactgatt	tggggatattt	gttggatgt	780
ttaaataaca	aacaagggtgg	tacctctggg	aaaactgact	ttggggaccat	ggctctaagg	840
ttgcatagaa	ggaaccattt	ttcgcctaga	aatacggatc	tttcccgga	tttggcaaaaa	900
gatcgtgtgg	ttatctgttt	gtatgtcat	aatcgggctc	agtattttcg	agtcacagt	960
gaaagtttgt	cgaagggttt	aggtaatagt	gagacattgt	tgttgttagt	tcatgtatgt	1020
tacttttgaag	agatgaatag	gattgtggag	agtttaatgt	tttgtcaagt	gaaacagatt	1080
ttctcgccctt	attcgcttca	tatatactgt	atagcttcc	ccgggtgtac	cctgaatgt	1140
tgtaaagaaca	aggggtatgt	ggcaaaagggg	cattgtgaag	gtaatcctga	tcagtatggg	1200
aatcatcggt	ctccgaagat	tgtatcttt	aagcatact	gttgggtggat	gatgaacact	1260
gtatggatgt	ggttggaaaga	gactaaagga	catgaggggc	atatcctttt	cattgaagaa	1320
gatcattttc	tttttcctaa	tgcctatctgt	aacatacaga	ctcttacag	gctgaaaccc	1380
gcaaaatgttc	ctgactgttt	tgctgtcat	ttagcaccgt	ctgtatgtaa	gtcaagagga	1440
gaagggcttgc	aaagttttgtt	tgccagagaga	atggggaaatg	ttgggttattt	ttttatata	1500
agtgtgtggg	agaatattca	tcagaaggca	agagatgtt	gttcccttta	tgattacaac	1560
tgggtatataa	cgatgtggc	aacgggtttt	ccgtcggtt	gttcccccgt	gtacacattt	1620
cgagggccta	ggactagtgc	ggtagacttt	ggaaaatgtt	gttgcacca	aggttagagga	1680
gatggggcttgc	attgcata	taatgggtt	gttaaacatag	aaatgtttagg	aacagataaa	1740
gttgtgaaca	taaaaagaagg	atggggagtt	cggtgtata	agcatcaagc	gggttataaa	1800
gcccgtttcg	agggtgggg	agggtggggc	gtatgtatgg	accgacattt	atgtttggat	1860
tttgcacta	tgtatgttca	cagcagtagc	agtgcata	catgaaacgg	atccgttca	1920
gtccgcaaaa	atcaccatgc	tctcttaca	aatctatctc	tctcttatttt	tctccagaat	1980
aatgtgtgag	tagttccac	ataagggaaat	taggttctt	atagggtttc	gctcatgtgt	2040
tgagcatata	agaaaccctt	agtatgtatt	tgtattttgt	aaataacttct	atcaataaaaa	2100
tttctaatcc	taaaaacccaaa	atcccgcgag	agaccttta	attaa		2145

<210> 42
<211> 541
<212> DNA
<213> Solanum tuberosum

<400>	42					
agatctcatt	ataccgttag	aagcatagg	aaaatctaaa	gcttgcgtt	aattctagtc	60
atttacatt	gttgggttct	acattattaa	tgaattttct	aatgcaaata	cagaattaa	120
atcaaaattg	ttgaattatg	ctaaacatgt	aacatacgt	tatctccgccc	tttgtgttgc	180
tattaacttg	aagtatcat	aagaaccaca	aatacactag	taaatctatg	agaaggcagg	240
tggcaacaca	aacaagaga	tctaagatt	tcatttgta	ctataggaat	ataatatctc	300
ttatctgatt	taatgaattcc	acatgttcac	ttctcatttg	tccacaagat	cacaacttta	360
tcttcataat	tcacaacttg	ttatatccac	cacaatttca	ttctttcact	ttagccccac	420
aaaatacttt	gtcccttat	ttgccaccc	ttgtattaa	tttattcttg	tggagctaag	480
tgttcatatt	attcttcttc	tcaaaaaaac	aaaaacaaaa	aaaaagagaa	aaaaaccatg	540
9						541

<210> 43
<211> 1358
<212> DNA
<213> hybrid

<400>	43	ccatggcgag	aggagcaga	tcagtggta	gcagcagcag	caaattggagg	tactgcaacc	60
cttccttata	cttgaaagcgc	ccaaagcgctc	ttgcgtctgt	cttcatcggtt	ttcggtttgtg			120
tctctttcgt	tttctggac	cgtcaaactc	tcgtcagaga	gcaccaggat	aaaatttctg			180
agctgcagaa	agaagtgact	gatttgaaaa	atttgggtgg	tgattnaat	aacaaacaag			240
gtggtacctc	tggaaaaact	gacttgggaa	ccatggctct	aagggttgcatt	agaaggaacc			300
attttcgccc	tagaaatacg	gatctgttcc	cgaggatttgc	aaaagatcg	gtgggttatcg			360
tcttgtatgt	gcataatcg	gctcaggatt	ttcgagtcac	agtggaaagt	ttgtcgaagg			420
ttaaaggat	aagttagag	tttgttatttgc	tttagtcatga	tggttacttt	gaagagatga			480
ataggattgt	ggagagttt	aagttttgtc	aagtggaaaca	gattttctcg	ccttattcgc			540
ctcatatata	tcgtactagc	ttccccgggt	tgaccctgaa	tgattgttaag	aacaagggtg			600
atgaggccaa	ggggcattgt	gaaggtaaatc	ctgatcagta	tggttacttt	cggtctccga			660
agattgtatc	tttgaagcat	cactgggtgt	ggatgtatgaa	cactgtatgg	gatgggttgg			720
aagagactaa	aggacatgag	gggcataatcc	ttttcatttgc	agaagatcat	tttctgtttc			780
ctaattgccta	tcgttaacata	cagactctta	cgaggctgaa	acccgc当地	tgtctgtact			840
gttttgcgtc	taatttttagca	ccgtctgtat	tgaagtcaag	aggagaagg	cttggaaagtt			900
tgggtgcaga	gagaatggg	aatgttgggt	atttttttaa	tagaagtgt	tgggagaata			960
ttcatcgagaa	ggcaagagag	ttttgttctt	ttgtatgatta	caactgggtat	ataacatgtt			1020
gggcacacgt	ttttccgtcg	ttttgttccc	cggtgtacac	attgcgaggg	ccttaggacta			1080
gtgcgtgtaca	cttggaaaaa	tgtgggttgc	atcaaggtag	aggagatgag	ggtgattgca			1140
tcgataatgg	ggtcgttaaac	atagaaggta	aggaaaacaga	taaagttgtg	aacataaaaag			1200
aaggatgggg	agttcggtgt	tataaggatc	aaggcggtt	taaagccgtt	ttcgaaggtt			1260
ggggaggttg	gggcgtatgt	agggaccgac	atttatgttt	ggattttgcc	actatgtatc			1320
qttacagcag	tagcagtgc	tctccatgaa	acggatcc					1358

<210> 44
<211> 237
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

```
<400> 44
ggatccgcta gagtccgcaa aatatcaccag tctctctcta caaatctatc tctctctatt 60
tttcccgaga ataatgtgtg agtagttccc agataaggga attagggttc ttatagggtt 120
tcgctcatgt gttgagcata taagaaaccc ttagtatgtt tttgtatttg taaaataactt 180
ctatcaataaa aatttctaat cctaaaaacca aaatccccgcg agagacctct taattaa 237
```

<210> 45
<211> 31
<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 45

atactcgagt taacaatgag taaacggaat c

31

<210> 46

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 46

ttctcgatcg ccgattggtt attc

24

<210> 47

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 47

gccgcccgcga tcgggcagtc ctcc

24

<210> 48

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 48

aacggatcca cgctagctcg gtgtcccgat

30

<210> 49

<211> 3327

<212> DNA

<213> hybrid

<400> 49

atgggcatca	agatggagac	acattctcg	gtctttgtat	acatgttgc	gtgggttgtct	60
ggtgtcgaca	tgaagcactt	caaattttcc	ctcactcaca	ccgtcaagag	ccgagacgag	120
ccaaacctccgg	atcaatgccc	tgcattgaag	gaaagcgaag	cggacatcga	caccgtggcg	180
atatacccaa	cttttgcattt	tcagccgagc	tggttgcgt	caaaggaaatt	ttgggacaag	240
tccttcgagg	atcggtatga	aagaattcat	aacgacacta	cacggccctag	actgaaggt	300
atcgtggttc	ctcaactcaca	caacgaccgc	ggatggctga	agacgtttga	acagtacttc	360
gagtggaaaga	ccaagaacat	tatcaacaac	atagtgaaca	aactgcacca	gtaccccaac	420
atgacccatca	tttggaccga	gatatcgctt	ctgatgcct	gtgtggaaag	gtcgcacccct	480
gtcaaaacaaa	aggcattgaa	aaaaacttatac	aaagaaggtc	gtctcgagat	cacgacgggc	540
ggctgggtga	tgcggacga	agcctgcacg	cataatctatg	cgctaattga	ccagtttatt	600
gaaggacatc	actgggtgaa	aactaatctc	ggcgtcatcc	cgaagacagg	atggttctatt	660
gaccccttcg	gccacggggc	cactgtgcct	tacctgctag	accagagcgg	ccttgaggga	720
accattatac	agagaatcca	ttatgcgtgg	aaacagtggc	tggcggagcg	acagattgag	780
gagtttact	ggctggcgg	ttgggtact	acgaagccgt	ccatgtatagt	gcacaatcag	840
ccgtttgtat	tttattcaat	aaaaaggcagc	tgtggccgc	acccttcaat	ttgtctcagt	900
ttcgactca	ggaagattcc	cggcgaatat	tctgaataca	cagctaagca	cgaagacatc	960
acggaaacaca	acttgcacag	caaggcataag	actttgtatag	aggagtacga	ccgtatcggg	1020

27/38

tccctgactc cacacaacgt ggtgctggg	1080
agcgtcgagt ttgatgccc atacgtcaat tatatgaaaa	1140
cacaaggaaa tcttcaacgc tgacgtacag ttcggaactc ctctcgatta cttaacgcc	1200
atgaaaagaaa gacatcaaaa tatacccgac ttaaaggag atttcttcgt ttactccgat	1260
atttcagcg aaggtaaacc agcgtactgg tcaggttact acactactag accctaccaa	1320
aaaatcctcg cccgtcgatt cgaacaccaa ctgcgatcgg cagagattt attcacccctt	1380
gtatcgaact acatcagaca gatgggtcgc caaggagagt tcggagcttc tgagaaaaag	1440
ttagaaaaat cttacgagca gcttatctat gtcgacgga acttgggtct gttcaacat	1500
cacgatgca ttacttggaa atcaaagtcc agtgtatgc aagattacgg aaccaaactg	1560
ttcacaagtc tgtatcactg catccgcctg caggaggccg cgctcaccac catcatgtt	1620
cctgaccagt cggtgcactc gcagagcatt atacaaagcg aggttgagtg ggaaacctac	1680
ggaaaacccgc ccaagaagct gcaagtgtcc ttcatgtaca agaagaaaagt tatactttt	1740
aatccgttgg ctgagactcg aactgaagtgt gtcacggta gatccaaacac gtccaaacatc	1800
cgggtgtacg atacacacaa gaggaaagcac gtcgttgcataat gataatgc cagcatcaca	1860
atccaagaca acggcaagag tatcgtaaac gacaccacgt tcgacataat gttcgtggcc	1920
accatcccgcc cccctcaccc tacatcgatc aagctcgagg agcacaaccaa cacttccac	1980
caactgcgtca ttttctgcaa caactgcgaa caataccaga aatccaatgt gttccaaatt	2040
aagaaaatga tgcctgtga catacaatta gaaaatgcag tgctaaaact tctcgtaat	2100
aggaacacccg gcttcttagt acaagtctat agaaaggaca tccggaagag aactgtcggt	2160
gacgtacaat tcggcgata tcaaagtgcc caaagacatt ctggtgctta cctcttcatg	2220
cctcattacg actcacctga gaagaatgtt ctgcatccct acactaatca gaacaacatg	2280
caagatgata acataatcat agtgtccgga cctatttcta cggaaatcac gaccatgtac	2340
ttgcctcttct tggtgacac tatttagata tacaacgtgc cggggccccgt actgtcggt	2400
gctattttat tagagaccga ttagatttc gagggcccac ctaagaacag agagactgag	2460
ttatatttga gattacagac tgatatacaa aacggtgaca ttcccgaatt ttacaccgt	2520
cagaacggat tccagtagca aaagagggtc aaagtgata aacttaggaat agaagactaat	2580
tactacccga tcactaccat ggcgtgcctg caagacgagg agacccggct cactctgtg	2640
acgaaccacg ctcaaggcgc tgctgcatac gaaccaggac gcttagaagt catgctcgat	2700
cgtcgaactc ttatgtatga cttagagggaa atcgggtgaag gagtagtgc taacaaacccg	2760
acgactttcc agaactggat ttaattgaa tccatgccc gcgtgacgcg agccaagaga	2820
gacactagtg aaccagggtt caaatttgtt aatgaacgtc gttttggccc cggccagaag	2880
gaaagccctt accaagtacc gtcgcagact gcggactacc tgagcaggat gttcaattac	2940
ccgggtgaacg tgtacctggt ggacacttagc gaggttggcg agatcgagggt gaagccgtac	3000
cagtcggtcc tgcagagctt cccgcccggc atccacctgg tcaccctgcg caccatcacc	3060
gacgacgtgc tcgaactctt cccccagcaac gaaagctaca tggtagtgc ccgaccagga	3120
tacagctgctg ctgtcgagaa gaagccagtc gccaagtctc ccaagtttc gtccaaaacc	3180
aggttcaatg gtcgaacat tcagaacatc actgcagtca gcctgaccgg cctgaagtca	3240
ctccgacctc tcacaggtct gagtgcacatc cacctgaacg ctatggaggt aaaaacttac	3300
aagatcaggt ttaaggacga gctttaa	3327

<210> 50

<211> 1108

<212> PRT

<213> hybrid

<400> 50

Met Gly Ile Lys Met Glu Thr His Ser Gln Val Phe Val Tyr Met Leu

1 5 10 15

Leu Trp Leu Ser Gly Val Asp Met Lys His Phe Lys Ser Ser Leu Thr

20 25 30

His Thr Val Lys Ser Arg Asp Glu Pro Thr Pro Asp Gln Cys Pro Ala

35 40 45

Leu Lys Glu Ser Glu Ala Asp Ile Asp Thr Val Ala Ile Tyr Pro Thr

50 55 60

Phe Asp Phe Gln Pro Ser Trp Leu Arg Thr Lys Glu Phe Trp Asp Lys

65 70 75 80

Ser Phe Glu Asp Arg Tyr Glu Arg Ile His Asn Asp Thr Thr Arg Pro

85 90 95

Arg Leu Lys Val Ile Val Val Pro His Ser His Asn Asp Pro Gly Trp

100 105 110

Leu Lys Thr Phe Glu Gln Tyr Phe Glu Trp Lys Thr Lys Asn Ile Ile
115 120 125

Asn Asn Ile Val Asn Lys Leu His Gln Tyr Pro Asn Met Thr Phe Ile
130 135 140

Trp Thr Glu Ile Ser Phe Leu Asn Ala Trp Trp Glu Arg Ser His Pro
145 150 155 160

Val Lys Gln Lys Ala Leu Lys Lys Leu Ile Lys Glu Gly Arg Leu Glu
165 170 175

Ile Thr Thr Gly Gly Trp Val Met Pro Asp Glu Ala Cys Thr His Ile
180 185 190

Tyr Ala Leu Ile Asp Gln Phe Ile Glu Gly His His Trp Val Lys Thr
195 200 205

Asn Leu Gly Val Ile Pro Lys Thr Gly Trp Ser Ile Asp Pro Phe Gly
210 215 220

His Gly Ala Thr Val Pro Tyr Leu Leu Asp Gln Ser Gly Leu Glu Gly
225 230 235 240

Thr Ile Ile Gln Arg Ile His Tyr Ala Trp Lys Gln Trp Leu Ala Glu
245 250 255

Arg Gln Ile Glu Glu Phe Tyr Trp Leu Ala Ser Trp Ala Thr Thr Lys
260 265 270

Pro Ser Met Ile Val His Asn Gln Pro Phe Asp Ile Tyr Ser Ile Lys
275 280 285

Ser Thr Cys Gly Pro His Pro Ser Ile Cys Leu Ser Phe Asp Phe Arg
290 295 300

Lys Ile Pro Gly Glu Tyr Ser Glu Tyr Thr Ala Lys His Glu Asp Ile
305 310 315 320

Thr Glu His Asn Leu His Ser Lys Ala Lys Thr Leu Ile Glu Glu Tyr
325 330 335

Asp Arg Ile Gly Ser Leu Thr Pro His Asn Val Val Leu Val Pro Leu
340 345 350

Gly Asp Asp Phe Arg Tyr Glu Tyr Ser Val Glu Phe Asp Ala Gln Tyr
355 360 365

Val Asn Tyr Met Lys Met Phe Asn Tyr Ile Asn Ala His Lys Glu Ile
370 375 380

Phe Asn Ala Asp Val Gln Phe Gly Thr Pro Leu Asp Tyr Phe Asn Ala
385 390 395 400

Met Lys Glu Arg His Gln Asn Ile Pro Ser Leu Lys Gly Asp Phe Phe
405 410 415

Val Tyr Ser Asp Ile Phe Ser Glu Gly Lys Pro Ala Tyr Trp Ser Gly
420 425 430

Tyr Tyr Thr Thr Arg Pro Tyr Gln Lys Ile Leu Ala Arg Gln Phe Glu
435 440 445

His Gln Leu Arg Ser Ala Glu Ile Leu Phe Thr Leu Val Ser Asn Tyr
450 455 460

Ile Arg Gln Met Gly Arg Gln Gly Glu Phe Gly Ala Ser Glu Lys Lys

29/38

465 470 475 480
Leu Glu Lys Ser Tyr Glu Gln Leu Ile Tyr Ala Arg Arg Asn Leu Gly
485 490 495
Leu Phe Gln His His Asp Ala Ile Thr Gly Thr Ser Lys Ser Ser Val
500 505 510
Met Gln Asp Tyr Gly Thr Lys Leu Phe Thr Ser Leu Tyr His Cys Ile
515 520 525
Arg Leu Gln Glu Ala Ala Leu Thr Thr Ile Met Leu Pro Asp Gln Ser
530 535 540
Leu His Ser Gln Ser Ile Ile Gln Ser Glu Val Glu Trp Glu Thr Tyr
545 550 555 560
Gly Lys Pro Pro Lys Lys Leu Gln Val Ser Phe Ile Asp Lys Lys Lys
565 570 575
Val Ile Leu Phe Asn Pro Leu Ala Glu Thr Arg Thr Glu Val Val Thr
580 585 590
Val Arg Ser Asn Thr Ser Asn Ile Arg Val Tyr Asp Thr His Lys Arg
595 600 605
Lys His Val Leu Tyr Gln Ile Met Pro Ser Ile Thr Ile Gln Asp Asn
610 615 620
Gly Lys Ser Ile Val Ser Asp Thr Thr Phe Asp Ile Met Phe Val Ala
625 630 635 640
Thr Ile Pro Pro Leu Thr Ser Ile Ser Tyr Lys Leu Gln Glu His Thr
645 650 655
Asn Thr Ser His His Cys Val Ile Phe Cys Asn Asn Cys Glu Gln Tyr
660 665 670
Gln Lys Ser Asn Val Phe Gln Ile Lys Lys Met Met Pro Gly Asp Ile
675 680 685
Gln Leu Glu Asn Ala Val Leu Lys Leu Leu Val Asn Arg Asn Thr Gly
690 695 700
Phe Leu Arg Gln Val Tyr Arg Lys Asp Ile Arg Lys Arg Thr Val Val
705 710 715 720
Asp Val Gln Phe Gly Ala Tyr Gln Ser Ala Gln Arg His Ser Gly Ala
725 730 735
Tyr Leu Phe Met Pro His Tyr Asp Ser Pro Glu Lys Asn Val Leu His
740 745 750
Pro Tyr Thr Asn Gln Asn Asn Met Gln Asp Asp Asn Ile Ile Val
755 760 765
Ser Gly Pro Ile Ser Thr Glu Ile Thr Thr Met Tyr Leu Pro Phe Leu
770 775 780
Val His Thr Ile Arg Ile Tyr Asn Val Pro Asp Pro Val Leu Ser Arg
785 790 795 800
Ala Ile Leu Leu Glu Thr Asp Val Asp Phe Glu Ala Pro Pro Lys Asn
805 810 815
Arg Glu Thr Glu Leu Phe Met Arg Leu Gln Thr Asp Ile Gln Asn Gly
820 825 830

Asp Ile Pro Glu Phe Tyr Thr Asp Gln Asn Gly Phe Gln Tyr Gln Lys
 835 840 845
 Arg Val Lys Val Asn Lys Leu Gly Ile Glu Ala Asn Tyr Tyr Pro Ile
 850 855 860
 Thr Thr Met Ala Cys Leu Gln Asp Glu Glu Thr Arg Leu Thr Leu Leu
 865 870 875 880
 Thr Asn His Ala Gln Gly Ala Ala Ala Tyr Glu Pro Gly Arg Leu Glu
 885 890 895
 Val Met Leu Asp Arg Arg Thr Leu Tyr Asp Asp Phe Arg Gly Ile Gly
 900 905 910
 Glu Gly Val Val Asp Asn Lys Pro Thr Thr Phe Gln Asn Trp Ile Leu
 915 920 925
 Ile Glu Ser Met Pro Gly Val Thr Arg Ala Lys Arg Asp Thr Ser Glu
 930 935 940
 Pro Gly Phe Lys Phe Val Asn Glu Arg Arg Phe Gly Pro Gly Gln Lys
 945 950 955 960
 Glu Ser Pro Tyr Gln Val Pro Ser Gln Thr Ala Asp Tyr Leu Ser Arg
 965 970 975
 Met Phe Asn Tyr Pro Val Asn Val Tyr Leu Val Asp Thr Ser Glu Val
 980 985 990
 Gly Glu Ile Glu Val Lys Pro Tyr Gln Ser Phe Leu Gln Ser Phe Pro
 995 1000 1005
 Pro Gly Ile His Leu Val Thr Leu Arg Thr Ile Thr Asp Asp Val
 1010 1015 1020
 Leu Glu Leu Phe Pro Ser Asn Glu Ser Tyr Met Val Leu His Arg
 1025 1030 1035
 Pro Gly Tyr Ser Cys Ala Val Gly Glu Lys Pro Val Ala Lys Ser
 1040 1045 1050
 Pro Lys Phe Ser Ser Lys Thr Arg Phe Asn Gly Leu Asn Ile Gln
 1055 1060 1065
 Asn Ile Thr Ala Val Ser Leu Thr Gly Leu Lys Ser Leu Arg Pro
 1070 1075 1080
 Leu Thr Gly Leu Ser Asp Ile His Leu Asn Ala Met Glu Val Lys
 1085 1090 1095
 Thr Tyr Lys Ile Arg Phe Lys Asp Glu Leu
 1100 1105

<210> 51
 <211> 1068
 <212> DNA
 <213> hybrid

<400> 51
 atgggcatca agatggagac acattctcag gtctttgtat acatgttgcgt gtgggtgtct 60
 ggtgtcgaca tgcagtccctc cggggagctc cggaccggag gggccccggcc gcccgcctcct 120
 ctaggcgcct cctcccgagcc gcgcgggggt ggcgactcca gcccagtcgt ggattctggc 180
 cctggccccc ctagcaactt gacccctggtc ccagtgcccc acaccacccgc actgtcgctg 240
 cccgcctgcc ctgaggagtc cccgtcgctt gtggggccca tgctgattga gtttaacatg 300
 cctgtggacc tggagctcggt ggcaaagcag aacccaaatg tgaagatggg cggccgctat 360
 gcccccaaggg actgcgtctc tcctcacaag gtggccatca tcattccatt ccgcacccgg 420

caggagcacc	tcaagtactg	gctatattat	ttgcacccag	tcctgcagcg	ccagcagctg	480
gactatggca	tctatgttat	caaccaggcg	ggagacacta	tattcaatcg	tgctaagctc	540
ctcaatgttg	gcttccaaga	agccttgaag	gactatgact	acacctgtct	tgtgtttagt	600
gacgtggacc	tcatccaat	aatgaccat	aatgcgtaca	ggtgttttc	acagccacgg	660
cacattttcg	ttgcaatggta	taagtttggta	ttcaggctac	ttatgttca	gtattttggta	720
ggtgtctctg	ctctaaagttaa	acaacagctt	ctaaacatca	atggatttcc	taataattat	780
tggggctggg	gaggagaaga	tgatgacatt	ttaaacatag	tagtttttag	aggcatgtct	840
atatctcgc	caaatgtctgt	ggtcggggagg	tgtcgcatga	ttcgccactc	aagagacaag	900
aaaaatgaac	ccaaatccctca	gaggtttgcac	cgaatgcac	acacaaaaggaa	gacaatgtct	960
tctgtatgtt	tgaactctact	caccttaccag	gtgctggatg	tacagagata	cccatttgtat	1020
acccaaatca	caqtggacat	cgggacaccg	agcaaggacg	agcttttag		1068

<210> 52
<211> 355
<212> PRT
<213> hybrid

<400> 52

Met Gly Ile Lys Met Glu Thr His Ser Gln Val Phe Val Tyr Met Leu
1 5 10 15

Leu Trp Leu Ser Gly Val Asp Met Gln Ser Ser Gly Glu Leu Arg Thr
20 25 30

Gly Gly Ala Arg Pro Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg
35 40 45

Pro Gly Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala
50 55 60

Ser Asn Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu
65 70 75 80

Pro Ala Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile
85 90 95

Glu Phe Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro
100 105 110

Asn Val Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro
115 120 125

His Lys Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu
130 135 140

Lys Tyr Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Gln Leu
345 150 155 160

Asp Tyr Gly Ile Tyr Val Ile Asn Gln Ala Gly Asp Thr Ile Phe Asn
165 170 175

Arg Ala Lys Leu Leu Asn Val Gly Phe Gin Glu Ala Leu Lys Asp Tyr
180 185 190

Asp Tyr Thr Cys Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met Asn
195 200 205

Asp His Asn Ala Tyr Arg Cys Phe Ser Gin Pro Arg His Ile Ser Val
210 215 220

Ala Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val Gin Tyr Phe Gly
225 230 235 240

Gly Val Ser Ala Leu Ser Lys Gin Gin Phe Leu Thr Ile Asn Gly Phe
245 250 255

32/38

Pro Asn Asn Tyr Trp Gly Trp Gly Gly Glu Asp Asp Asp Ile Phe Asn
 260 265 270

Arg Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val
 275 280 285

Gly Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Lys Asn Glu Pro
 290 295 300

Asn Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met Leu
 305 310 315 320

Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg
 325 330 335

Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser Lys
 340 345 350

Asp Glu Leu
 355

<210> 53
<211> 1119
<212> DNA
<213> hybrid

<400> 53

atggggcatca	agatggagac	acattctcg	gtctttgtat	acatgttgct	gtgggtgtct	60
ggtgtcaca	tggacagat	gcctgtggct	gctgttagtgg	ttatggcctg	cagtctgca	120
gactatctt	aaaggactgt	taaatcagtt	ttaacatatc	aaactcccg	tgcttcaaaa	180
tatcctctat	ttatatctca	ggatggatct	gatcaagctg	tcaagagcaa	gtcattgagc	240
tataatcaat	taacatataat	gcagcactt	gatttgaac	cagtggtcac	tgaaaggcct	300
ggcgaactga	ctgcgtacta	caagattgca	cgtcaactaca	agtgggcact	ggaccagtt	360
ttttacaaaac	acaaaatttag	tcgagtgatt	atactagaag	atgatatgga	aattgtcca	420
gacttctt	attactttga	ggctgcagct	agtctcatgg	atagggataaa	aaccattatg	480
gctgcttcat	catggaatga	taatggacag	aagcagttt	tgcatgatcc	ctatgcgcta	540
taccgtatcg	attttttcc	tggccttggg	tggatgtca	agagatcgac	ttgggatgag	600
ttatcaccaa	agtggccaaa	ggcttactgg	gatgattggc	tgagatctaa	ggaaaaccat	660
aaaggccgccc	aattcattcg	accggaagtc	tgtagaacat	acaattttgg	tgaacatggg	720
tctagtttgg	gacagtttt	cagtcagat	ctggAACCTA	taaagctaaa	cgatgtgacg	780
gttgacttgg	aaagcaaagga	cctgggatatac	ctgacagagg	aaaactatac	caagtacttt	840
tctggcttag	tgagacaagc	acgaccaatt	caaggttctg	accttgcctt	aaaggctcaa	900
aacataaaagg	atgatgttcg	tatccgtat	aaagaccaag	tagagttga	acgcattgca	960
ggggaaatttg	gtatatttga	agaatggaaag	gatgtgtgc	ctcgaacagc	atataaagga	1020
gtagtgggt	ttcgaatcca	gacaacaaga	cgtgtattcc	tggttgggccc	agattctgta	1080
atgcagctt	gaattcggaaa	ttccaaggac	gagcttga			1119

<210> 54
<211> 372
<212> PRT
<213> hybrid

<400> 54

Met Gly Ile Lys Met Glu Thr His Ser Gln Val Phe Val Tyr Met Leu
 1 5 10 15

Leu Trp Leu Ser Gly Val Asp Met Gly Gln Met Pro Val Ala Ala Val
 20 25 30

Val Val Met Ala Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys
 35 40 45

Ser Val Leu Thr Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe
 50 55 60

Ile Ser Gln Asp Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser
 65 70 75 80

Tyr Asn Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val
 85 90 95

Thr Glu Arg Pro Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His
 100 105 110

Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg
 115 120 125

Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp
 130 135 140

Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met
 145 150 155 160

Ala Ala Ser Ser Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp
 165 170 175

Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met
 180 185 190

Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala
 195 200 205

Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln
 210 215 220

Phe Ile Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly
 225 230 235 240

Ser Ser Leu Gly Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu
 245 250 255

Asn Asp Val Thr Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr
 260 265 270

Glu Gly Asn Tyr Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg
 275 280 285

Pro Ile Gln Gly Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp
 290 295 300

Asp Val Arg Ile Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala
 305 310 315 320

Gly Glu Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr
 325 330 335

Ala Tyr Lys Gly Val Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val
 340 345 350

Phe Leu Val Gly Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser
 355 360 365

Lys Asp Glu Leu
 370

<210> 55
 <211> 1158
 <212> DNA
 <213> hybrid

<400> 55
 atgggcatca agatggagac acattcttag gtctttgtat acatgttgcgt gtggttgtct 60

ggtgtcgaca	tggctctaag	gttgcataga	aggaaccatt	tttcgcctag	aaatacggat	120
ctgttcccgg	attggcaaa	agatcgtgtg	tttatcgct	tgtatgtgc	taatcggt	180
cagtatttc	gagtcacagt	ggaaaagttt	tcgaaggta	aaggtaataag	tgagacattg	240
ttgattgtta	gtcatgatgg	ttactttgaa	gagatgaata	ggattgtgga	gagttataag	300
ttttgtcaag	tgaaacagat	tttctcgcc	tattcgcc	atatatatcg	tactagctc	360
ccgggtgtga	ccctgaatga	ttgtaaagaac	aagggtgtg	aggcaaagg	gcattgtgaa	420
ggttaatcctg	atcagtatgg	gaatcatcg	tctccgaaga	ttgtatctt	gaagcatcac	480
tggtgggtga	tgatgaacac	tgtatgggat	gggttggaaag	agactaaagg	acatgagggg	540
catatccctt	tcattgaaga	agatcattt	ctgttccct	atgcctatcg	taacatacag	600
actcttacga	ggctgaaacc	cgcaaaagt	cctgactgtt	ttgctgctaa	tttagcaccg	660
tctgatgtga	agtcaagagg	agaagggtt	gaaagtttgg	ttgcagagag	aatggaaat	720
gttgggtatt	ctttaatag	aagtgtgtt	gagaatattc	atcagaaggc	aagagat	780
tgttcttt	atgattacaa	ctggatata	acgtgtggg	caacggttt	ccgtcg	840
ggttccccgg	tgtacacatt	gcgaggcct	aggactagt	cggtaacatt	tggaaatgt	900
gggttgcac	aagtagagg	agatgagggt	gattgcatcg	ataatgggg	cgtaaacata	960
gaagtttaagg	aaacagataa	agttgtgaac	ataaaaagaag	gatggggagt	tcgggtgtat	1020
aagcatcaag	cgggttataa	agccggttt	gaagttggg	gaggttgggg	cgatgatagg	1080
gaccgacatt	tatgtttgga	tttgcact	atgtatcg	acagcagtag	cagtgcac	1140
ccaaaggacg	agctttga					1158

<210> 5b

<211> 385

<212> PRT

<213> hybrid

<400> 5b

Met	Gly	Ile	Lys	Met	Glu	Thr	His	Ser	Gln	Val	Phe	Val	Tyr	Met	Leu
1				5					10					15	

Leu	Trp	Leu	Ser	Gly	Val	Asp	Met	Ala	Leu	Arg	Leu	His	Arg	Arg	Asn
					20				25					30	

His	Phe	Ser	Pro	Arg	Asn	Thr	Asp	Leu	Phe	Pro	Asp	Leu	Ala	Lys	Asp
					35			40				45			

Arg	Val	Val	Ile	Val	Leu	Tyr	Val	His	Asn	Arg	Ala	Gln	Tyr	Phe	Arg
					50			55			60				

Val	Thr	Val	Glu	Ser	Leu	Ser	Lys	Val	Lys	Gly	Ile	Ser	Glu	Thr	Leu
					65			70			75			80	

Leu	Ile	Val	Ser	His	Asp	Gly	Tyr	Phe	Glu	Glu	Met	Asn	Arg	Ile	Val
					85			90			95				

Glu	Ser	Ile	Lys	Phe	Cys	Gln	Val	Lys	Gln	Ile	Phe	Ser	Pro	Tyr	Ser
					100			105			110				

Pro	His	Ile	Tyr	Arg	Thr	Ser	Phe	Pro	Gly	Val	Thr	Leu	Asn	Asp	Cys
					115			120			125				

Lys	Asn	Lys	Gly	Asp	Glu	Ala	Lys	Gly	His	Cys	Glu	Gly	Asn	Pro	Asp
					130			135			140				

Gln	Tyr	Gly	Asn	His	Arg	Ser	Pro	Lys	Ile	Val	Ser	Leu	Lys	His	His
					145			150			155			160	

Trp	Trp	Trp	Met	Met	Asn	Thr	Val	Trp	Asp	Gly	Leu	Glu	Glu	Thr	Lys
					165			170			175				

Gly	His	Glu	Gly	His	Ile	Leu	Phe	Ile	Glu	Glu	Asp	His	Phe	Leu	Phe
					180			185			190				

Pro	Asn	Ala	Tyr	Arg	Asn	Ile	Gln	Thr	Leu	Thr	Arg	Leu	Lys	Pro	Ala
					195			200			205				

Lys	Cys	Pro	Asp	Cys	Phe	Ala	Ala	Asn	Leu	Ala	Pro	Ser	Asp	Val	Lys
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

210	215	220
Ser Arg Gly Glu Gly Leu Glu Ser Leu Val Ala Glu Arg Met Gly Asn		
225	230	235
240		
Val Gly Tyr Ser Phe Asn Arg Ser Val Trp Glu Asn Ile His Gln Lys		
245	250	255
Ala Arg Glu Phe Cys Phe Phe Asp Asp Tyr Asn Trp Asp Ile Thr Met		
260	265	270
Trp Ala Thr Val Phe Pro Ser Phe Gly Ser Pro Val Tyr Thr Leu Arg		
275	280	285
Gly Pro Arg Thr Ser Ala Val His Phe Gly Lys Cys Gly Leu His Gln		
290	295	300
Gly Arg Gly Asp Glu Gly Asp Cys Ile Asp Asn Gly Val Val Asn Ile		
305	310	315
320		
Glu Val Lys Glu Thr Asp Lys Val Val Asn Ile Lys Glu Gly Trp Gly		
325	330	335
Val Arg Val Tyr Lys His Gln Ala Gly Tyr Lys Ala Gly Phe Glu Gly		
340	345	350
Trp Gly Gly Trp Gly Asp Asp Arg Asp Arg His Leu Cys Leu Asp Phe		
355	360	365
Ala Thr Met Tyr Arg Tyr Ser Ser Ser Ser Ala Ser Pro Lys Asp Glu		
370	375	380
Leu		
385		

<210> 57
<211> 1152
<212> DNA
<213> Homo sapiens

<400> 57
atgctgaaga agcagtctgc agggcttgc ctgtggcg ctatccttt tggccctgg 60
aatgccttc tgctccttt cttctggacg cgccagcac ctggcaggcc accctcagtc 120
agcgcttcg atggcacc ccgcagcctc acccgaaag tcgacatgca gtcctccgg 180
gagctccgga ccggaggggc ccggccgcgg cctccttag ggcctccctc ccagccgcgc 240
ccgggtggcg actccagccc agtcgtggat tctggccctg gccccgttag caacttgacc 300
tcggctcccg tgccccacac caccgcactg tcgctgccc cctgcccctg ggagtccccg 360
tcgcttgtgg gccccatgt gattgagtt aacatgcctg tggacctgg gtcgtggca 420
aagcagaacc caaatgtgaa gatgggcggc cgctatgccc ccaggactg cgtcttcct 480
cacaagggtgg ccatcatcat tccattccgc aacccgcagg agcacctcaa gtactggcta 540
tattatttgc acccagtctt gcagcgccag cagctggact atggcatcta tgttatcaac 600
caggcgggag acactatatt caatcggt aagctcctca atgttggctt tcaagaagcc 660
ttgaaggact atgactacac ctgctttgtg tttagtgcg tggacctcat tccaatgaat 720
gaccataatg cgtacaggtt ttttcacag ccacggcaca tttccgttgc aatggataag 780
tttggattca gcctaccta tgttcagat tttggagggtg tctctgtct aagtaaaca 840
cagtttctaa ccatcaatgg atttcctaatttatttgg gctggggagg agaagatgtat 900
gacattttta acagattagt ttttagaggc atgtctatat ctgcggccaaa tgctgtggc 960
gggagggtgtc gcatgatccg ccactcaaga gacaagaaaa atgaaccccaa tcctcagagg 1020
tttggccaa ttgcacacac aaaggagaca atgcctctg atggtttggaa ctcactcacc 1080
taccagggtgc tggatgtaca gagataccca ttgtataccca aatcacagt ggacatcg 1140
acaccggact ag 1152

<210> 58
<211> 383
<212> PRT
<213> Homo sapiens

<400> 58

Met Leu Lys Lys Gln Ser Ala Gly Leu Val Leu Trp Gly Ala Ile Leu
1 5 10 15

Phe Val Ala Trp Asn Ala Leu Leu Leu Phe Phe Trp Thr Arg Pro
20 25 30

Ala Pro Gly Arg Pro Pro Ser Val Ser Ala Leu Asp Gly Asp Pro Ala
35 40 45

Ser Leu Thr Arg Glu Val Asp Met Gln Ser Ser Gly Glu Leu Arg Thr
50 55 60

Gly Gly Ala Arg Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg
65 70 75 80

Pro Gly Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala
85 90 95

Ser Asn Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu
100 105 110

Pro Ala Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile
115 120 125

Glu Phe Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro
130 135 140

Asn Val Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro
145 150 155 160

His Lys Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu
165 170 175

Lys Tyr Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Gln Leu
180 185 190

Asp Tyr Gly Ile Tyr Val Ile Asn Gln Ala Gly Asp Thr Ile Phe Asn
195 200 205

Arg Ala Lys Leu Leu Asn Val Gly Phe Gln Glu Ala Leu Lys Asp Tyr
210 215 220

Asp Tyr Thr Cys Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met Asn
225 230 235 240

Asp His Asn Ala Tyr Arg Cys Phe Ser Gln Pro Arg His Ile Ser Val
245 250 255

Ala Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val Gln Tyr Phe Gly
260 265 270

Gly Val Ser Ala Leu Ser Lys Gln Gln Phe Leu Thr Ile Asn Gly Phe
275 280 285

Pro Asn Asn Tyr Trp Gly Trp Gly Gly Glu Asp Asp Asp Ile Phe Asn
290 295 300

Arg Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val
305 310 315 320

Gly Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Lys Asn Glu Pro
325 330 335

Asn Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met Leu
340 345 350

Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg
355 360 365

Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser
370 375 380

<210> 59
<211> 400
<212> PRT
<213> Homo sapiens

<400> 59

Met Arg Leu Arg Glu Pro Leu Leu Ser Gly Ala Ala Met Pro Gly Ala
1 5 10 15

Ser Leu Gln Arg Ala Cys Arg Leu Leu Val Ala Val Cys Ala Leu His
20 25 30

Leu Gly Val Thr Leu Val Tyr Tyr Leu Ala Gly Arg Asp Leu Ser Arg
35 40 45

Leu Pro Gln Leu Val Gly Val Ser Thr Pro Leu Gln Gly Gly Ser Asn
50 55 60

Ser Ala Ala Ala Ile Gly Gln Ser Ser Gly Glu Leu Arg Thr Gly Gly
65 70 75 80

Ala Arg Pro Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg Pro Gly
85 90 95

Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala Ser Asn
100 105 110

Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu Pro Ala
115 120 125

Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile Glu Phe
130 135 140

Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro Asn Val
145 150 155 160

Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro His Lys
165 170 175

Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu Lys Tyr
180 185 190

Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Gln Leu Asp Tyr
195 200 205

Gly Ile Tyr Gly Ile Tyr Val Ile Asn Gln Ala Gly Asp Thr Ile Phe
210 215 220

Asn Arg Ala Lys Leu Leu Asn Val Gly Phe Gln Glu Ala Leu Lys Asp
225 230 235 240

Tyr Asp Tyr Thr Cys Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met
245 250 255

Asn Asp His Asn Ala Tyr Arg Cys Phe Ser Gln Pro Arg His Ile Ser
260 265 270

Val Ala Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val Gln Tyr Phe
275 280 285

Gly Gly Val Ser Ala Leu Ser Lys Gln Gln Phe Leu Thr Ile Asn Gly
290 295 300

Phe Pro Asn Asn Tyr Trp Gly Trp Gly Gly Glu Asp Asp Asp Ile Phe
305 310 315 320

Asn Arg Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val
325 330 335

Val Gly Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Lys Asn Glu
340 345 350

Pro Asn Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met
355 360 365

Leu Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln
370 375 380

Arg Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser
385 390 395 400